התאמות חלקיות של פרופילים גנטיים באוכלוסייה רחבה: תיאוריה, תחזיות, דיון משפטי ומשמעויות מעשיות

מרדכי הלפרט*, יורם פלוצקי ודוד אואן*****

"The argument that the same risk was flown before without failure is often accepted as an argument for the safety of accepting it again. Because of this, obvious weaknesses are accepted again and again, sometimes without a sufficiently serious attempt to remedy them, or to delay a flight because of their continued presence."[1]

תקציר

בצו בית משפט בארה"ב, בוצעה במדינת אריזונה ספירת ההתאמות החלקיות בין פרופילים גנטיים במאגר DNA שהכיל 65,493 פרופילים, שהופקו ב-13 אתרים. נמצא כי 122 זוגות פרופילים גנטיים תואמים ביניהם ב-9 אתרים מתוך ה-13, 20 זוגות פרופילים תואמים ב-10 אתרים מה-13, זוג אחד תואם ב-11 אתרים וזוג נוסף תואם ב-12 אתרים מתוך ה-13. בסה"כ נמצאו 144 התאמות חלקיות ב-9 אתרים ויותר. נסקרת הספרות שדנה בתוצאות שהתקבלו באריזונה וההסברים המוצעים לממצאים.

במאמר הנוכחי נבחנות השלכות הנושא על המשפט בישראל. מתואר מודל תיאורטי על פיו ניתן לְחַשב את מספר ההתאמות החלקיות הצפוי במאגר ובאוכלוסייה כולה, על סמך עקרונות גנטיים מקובלים. המודל מביא בחשבון גם את האפשרות כי האוכלוסייה אינה הומוגנית אלא מורכבת ממספר אוכלוסיות ייחוס. המודל כולל תיקון θ (טטה) לתת-אוכלוסיות בכל אוכלוסיית ייחוס. המודל אינו מביא בחשבון שגיאות מעבדה והתאמות בין קרובי משפחה. לבדיקת תקפות המודל נערכת סימולציה ממוחשבת בה מוגרלים פרופילים אקראיים על פי טבלאות שכיחות אללים באוכלוסייה. תוצאות הסימולציה תואמות את התוצאות מהמודל התיאורטי עבור פרופילים גנטיים בישראל המופקים בעשרה אתרים.

בהסתמך על המודל התיאורטי, מוגשת תחזית למספר ההתאמות החלקיות הצפוי במאגר גנטי בישראל, כפונקציה של גודלו, עבור פרופילים גנטיים המופקים בעשרה אתרים בעזרת הערכה SGM PLUS. השוואה בין תוצאות התאמות בפועל מן המאגר בישראל לבין החזוי על פי המודל, יכולה לשמש מבחן הפרכה לתיקון טטה המוצג על ידי מומחי התביעה בבתי משפט בישראל. מובאת גם תחזית כזאת עבור כל האוכלוסייה. נדונות המשמעויות המשפטיות הכלליות הנובעות מעצם קיום התאמות חלקיות במספר גדול של אתרים. נדונות גם המשמעויות המעשיות של התאמות חלקיות לתיקי עבודה, במקרה של הפקת פרופיל חלקי בלבד מזירת הפשע, ובמקרה בו מתקבלים פרופילים שגויים בבדיקות. מודגש השילוב המסוכן בין התאמה חלקית לבין התופעה הנפוצה של allelic dropout שעלולה להוביל לזיהוי שגוי. מובאות מספר המלצות לצמצום אפשרות של זיהוי שגוי. מוצע למערכת המשפט שלא לבסס הרשעה על רְאָיַת DNA כרְאָיָה יחידה.

* ד"ר מרדכי הלפרט, Ph.D, פיסיקאי העוסק במחקר ופיתוח בתעשייה mhalpert@zahav.net.il
** ד"ר יורם פלוצקי Ph.D בגנטיקה, עו"ד – מרצה מן החוץ באוניברסיטת בר אילן ומנכ"ל חברת GALIL GENETIC ANALYSIS (GGA)
*** פרופסור (אמריטוס) דוד אואן מחלקה לפיסיקה באוניברסיטת בן גוריון בנגב.

1 Richard P. Feynman, Richard Feynman's Minority Report to the Space Shuttle Challenger Inquiry, in RICHARD P. FEYNMAN, THE PLEASURE OF FINDING THINGS OUT: THE BEST SHORT WORKS OF RICHARD P. FEYNMAN, 155 (Jeffrey Robbins ed., 1999), available at: http://science.ksc.nasa.gov/shuttle/missions/51-l/docs/rogers-commission/Appendix-F.txt (Kennedy Space Center website).
ראו בתרגום לעברית, ריצ'רד פיינמן "דוח המיעוט של ריצ'רד פ' פיינמן לחקירת מעבורת החלל צ'לנג'ר" ריצ'רד פיינמן, **החדווה שבגילוי דברים**, 163. (כנרת, זמורה ביתן, דביר 2006).

א. הקדמה

התאמה חלקית בין פרופילים גנטיים, בהקשר של מאמר זה, היא מצב בו פרופיל גנטי תואם לפרופיל אחר במספר חלקי של אתרים, והוא אינו תואם בשאר האתרים המרכיבים את הפרופיל.[2] פרופיל חלקי בהקשר של מאמר זה הוא פרופיל שכתוצאה מבעיות טכניות, הופק רק בחלק מן האתרים. התאמה חלקית באתר, בהקשר של מאמר זה היא מצב בו שני פרופילים גנטיים זהים באלל אחד מתוך שנים שבאותו האתר.

על פי צו של בית משפט אמריקני, בוצעה באריזונה ספירת התאמות חלקיות במאגר פרופילים גנטיים.[3] התוצאות הפתיעו רבים. המאגר מנה 65,493 פרופילים גנטיים בני 13 אתרים. נמצאו בו 122 זוגות פרופילים גנטיים התואמים ב-9 אתרים. 22 זוגות פרופילים גנטיים התואמים ב-10 אתרים, זוג אחד של פרופילים התואם ב-11 אתרים וזוג נוסף התואם ב-12 אתרים מתוך ה-13. בסה"כ, 144 זוגות פרופילים גנטיים התואמים בתשעה אתרים ויותר.[4] ההפתעה נובעת מכך שהסתברות התאמה המחושבת באותם 9 ויותר אתרים בהם נמצאה התאמה בדרך המקובלת, היא בסדר גודל של אחד למיליארדים ואף פחות מכך (עבור התאמה ב-13 אתרים מדובר על הסתברויות התאמה של מקרה אחד לטריליונים רבים). בכל זאת, כבר בקרב עשרות אלפי פרופילים גנטיים נמצא מספר גדול יחסית (מאות) של זוגות פרופילים תואמים במספר גדול של אתרים. התוצאות שעלו באריזונה מלמדות כי לאחד מבין 227 פרופילים גנטיים, יימצא תואם בתשעה אתרים ויותר, אף בקרב 65,493 פרופילים גנטיים.[5]

בשנים האחרונות הוחל בישראל בהפקת פרופילים גנטיים בעשרה אתרים בעזרת העֶרְכָּה SGM PLUS (kit) המיוצרת על ידי חברת Applied Biosystems. מאמר זה בוחן את ההשלכות של ההתאמות החלקיות הצפויות במאגר פרופילים ובאוכלוסייה כולה, על השימוש ברְאָיַת ה-DNA בישראל. בפרק ב' נסקור את הספרות הדנה בתוצאות שהתקבלו במאגר באריזונה. בפרק ג' נתאר מודל תיאורטי המבוסס על ההנחות הגנטיות המקובלות, המְחַשב את מספר ההתאמות החלקיות הצפוי במאגר. המודל התיאורטי מביא בחשבון שמקור התורמים למאגר הוא מכמה קבוצות אוכלוסייה. המודל כולל בתוכו גם את תיקון θ ("טטה") עבור תת אוכלוסיות.[6] המודל איננו מביא בחשבון שגיאות מעבדה וקיומם של אחים מלאים במאגר. הפרטים הטכניים של המודל מופיעים בפירוט רב בנספח הטכני למאמר זה. בפרק ד' נבחן את מספר ההתאמות שהמודל מנבא למול סימולציה ממוחשבת של מאגר פרופילים בישראל. מטרת הסימולציה היא לבדוק את המודל התיאורטי ולתת לו תוקף במסגרת ההנחות עליהן הוא מבוסס. התוצאות העולות מהסימולציה שביצענו תואמות את התוצאות מהמודל התיאורטי. בפרק ה', ניתן תחזית תיאורטית למספר ההתאמות החלקיות הצפוי במאגר גנטי בישראל כפונקציה של גודלו. בפרק ו' ניתן תחזית כזאת עבור האוכלוסייה הכללית בישראל. בפרק ז', נדון במשמעות המשפטית הכללית של התוצאות. בפרק ח' נדון בהשלכות מעשיות להתאמות חלקיות שיכולות להתעורר בתיקי עבודה בבתי משפט בישראל, לאור אפשרות תופעות שונות המובילות להפקת פרופיל שגוי ולאור האפשרות כי מזירת פשע יופק פרופיל חלקי בלבד. המאמר יתמקד בשילוב המסוכן של התאמות חלקיות עם התופעה הנפוצה של allelic dropout- העלול להוביל לזיהוי שגוי.

ב. התוצאות שהתקבלו במאגר באריזונה

העובדה שבמאגר המכיל עשרות אלפי פרופילים גנטיים התגלו התאמות חלקיות במספר גדול של אתרים איננה צריכה להפתיע.[7] הסיבה לכך היא כי איננו משווים פרופיל גנטי אחד לכלל הפרופילים הגנטיים אלא אנו משווים את כל

2 בדיקת ה-DNA מתוארת רבות ובפירוט רב בספרות. להעמקה ראו אחיקם סטולר ויורם פלוצקי "D.N.A. על דוכן העדים" **רפואה ומשפט** 25, 143 (2001) ; נירה גלילי ואסא מרבך "אנליזה של דנ"א למטרות פורנזיות" **פלילים** ב 225 (1991) ; מרדכי הלפרט ומשה פרדס, "האומנם ניתן להרשיע על בסיס ראיה מדעית יחידה? המקרה של ראיות טביעת האצבע ו-DNA", **עיוני משפט** ל 399 (2007) ; Fed Judicial Ctr., Reference Manual on Scientific Evidence 487–576 (2nd ed., 2000); B. Sangero and M. Halpert, Why a Conviction Should Not Be Based on a Single Piece of Evidence: A Proposal for Reform, 48 Jurimetrics J. 43, 72 (2007);

3 Myers S. P. 2006 Felon-to-felon STR partial profile matches in the Arizona database: don't panic. Presentation at the California association of criminalists DNA workshop.

4 Laurence D. Mueller, Can simple population genetic models reconcile partial match frequencies observed in large forensic databases? 87 J Genet. 101 (Aug 2008).

5 144 זוגות פרופילים גנטיים הם 288 פרופילים גנטיים. $65493/288 = 227.4$

6 ראה הסבר לתיקון טטה בפרק ג' וכן בסעיף 6 לנספח הטכני.

7 עבור פרופילים גנטיים המורכבים מתשעה אתרים, כפי שהיה נהוג בבדיקות בישראל עד לפני מספר לא רב של שנים, נצפה להתאמה מלאה

הפרופילים בינם לבין עצמם. אם המאגר מונה N פרופילים גנטיים אזי מספר השוואות יהיה $\frac{N\times(N-1)}{2}$.[8] עבור המאגר באריזונה הדבר מוביל ללמעלה מ-2 מיליארד השוואות. מכאן, שלא תהה זאת הפתעה גדולה אם באותן 2 מיליארד השוואות ימצאו התאמות חלקיות גם כאשר הסתברות ההתאמה חלקית של הפרופילים במאגר היא בסדרי גודל של אחד למיליארדים ופחות.

יתרה מכך, ההתאמות החלקיות שנמצאו במאגר אינן מתרחשות במספר אתרים קבוע שנקבע מראש. כך למשל, יש 715 אפשרויות לבחור 9 אתרים מתוך 13.[9] מכאן, שכאשר מבצעים השוואה לבחינת התאמה חלקית בתשעה אתרים, לא מבצעים השוואה אחת אלא 715 השוואות שכאלו. הדבר מעלה כמובן את הסתברות ההתאמה. מכאן שהשאלה המדעית איננה על עצם קיומן של התאמות חלקיות רבות במספר רב של פרופילים גנטיים במאגר, אלא האם מספרם תואם את התיאוריה הגנטית.

לאחרונה, פורסם מחקר של Mueller, בו הוא מנתח את תוצאות הממצאים שבמאגר אריזונה.[10] המחקר בוצע בעזרת סימולציה ממוחשבת של פרופילים גנטיים. Mueller הניח כי המאגר כולל פרופילים גנטיים מתורמים השייכים לאוכלוסיות שונות. מאחר ולא היו בידו נתונים לגבי אוכלוסיית התורמים למאגר, הוא הניח כי התפלגות האוכלוסיות התורמות למאגר זהות להתפלגות אוכלוסיות האסירים שבבתי הסוהר עליהם פורסמו נתונים.[11] האוכלוסיות שהובאו בחשבון הן African American, Caucasian, Hispanic, Navajo, Apache.[12] הנחות נוספות היו שאין בין אלה תת-אוכלוסיות וגם אין ביניהן קרובי משפחה. בסימולציה נספרו 100 זוגות פרופילים התואמים חלקית בתשעה אתרים ו-3.9 פרופילים התואמים ב-10 אתרים. חשיבות המחקר של Mueller היא באספקת מודל שגיאה למספר ההתאמות הצפויות במאגר ובכך הוא בוחן אם התיאוריה מתאימה לממצאים בפועל שבמאגר. התוצאות שהתקבלו בפועל במאגר באריזונה נופלות מחוץ לאליפסת בטחון (confidence ellipsoids) של 99% מן החזוי בסימולציה. הבאה בחשבון של תת-אוכלוסיות בלבד, או אחים בלבד, לא הובילה להתאמה בין ממצאי הסימולציה לתוצאות שהתקבלו בפועל, ורק הגדילה את מספר ההתאמות הצפוי בתשעה אתרים, מעבר ל-122 ההתאמות הידועות, ללא הגדלה מספקת של מספר ההתאמות בעשרה אתרים. רק הבאה בחשבון של פרופילים המשתייכים ל-4 תת-אוכלוסיות, בכל אוכלוסיית ייחוס, עבור טווח ערכים מסוים של θ, עם הכנסת טווח מסוים של מספר קרובי משפחה בדרגת אחים מלאים, הביאה להתאמה באליפסת בטחון של 95% בין תוצאות הסימולציה לבין הממצאים בפועל. מאחר והתיאוריה מתאימה לממצאים בפועל רק בטווח ערכים מאד מסוים של תת-אוכלוסיות וטווח ערכים מצומצם של תורמים קרובי משפחה בדרגת אחים מלאים, מסיק Mueller כי לבחינת סבירות התוצאות ותקפות הנחות העבודה הגנטיות, יש צורך שהרָשויות יספקו מידע נוסף על אותו מאגר.

Weir שחקר את התוצאות באריזונה,[13] הציע מודל תיאורטי, אשר הנחת היסוד שבו היא שהמאגר מורכב מאוכלוסיית ייחוס אחת בלבד (Caucasian). החישוב שלו נעשה כאשר ערך $\theta = 0.03$. הוא גם מחשב את מספר ההתאמות החלקיות הצפוי באתרים בהם לא נמצאה התאמה מלאה בשני אללים. תופעה זאת לא נבדקה במאגר באריזונה ולכן לא ניתן להשוות את תוצאת החישוב עם מה שנמצא בפועל במאגר. התוצאות שלו[14] אכן מנבאות 20 התאמות מלאות (שני אללים זהים) ב-10 אתרים מתוך 13, כאשר מתוך אותם 20, 3 זוגות פרופילים תואמים חלקית גם בשלושת האתרים הנוספים (23 אללים זהים מתוך 26), 8 תואמים חלקית ב-2 מתוך 3 האתרים הנוספים, 7 תואמים חלקית באתר אחד מבין 3 האתרים הנוספים ו-2 זוגות לא תואמים חלקית באף אחד משלושת האתרים הנוספים.[15] מכאן כי על פי התחזית

כבר בקרב 57,413 פרופילים גנטיים. ראו, מרדכי הלפרט, על ייחודיות פרופיל ה-DNA, תיאוריה ותוצאות מסימולציה ממוחשבת של פרופילים גנטיים באוכלוסייה רחבה., **רפואה ומשפט** 39, עמ' 97 (דצמבר 2008). על השלכות העניין עבור מקרים בעבר בהם הורשעו אנשים על סמך התאמה בשישה אתרים, ראו יורם פלוצקי, הערות למאמר על ייחודיות פרופיל ה-DNA, תיאוריה ותוצאות מסימולציה ממוחשבת של פרופילים גנטיים באוכלוסייה רחבה. 39, **רפואה ומשפט** 115, 116 (דצמבר 2008).

8 הלפרט, לעיל ה"ש 7, בעמ' 114.

9 מספר האפשרויות לבחור $w=9$ אתרים מתוך $n=13$ אתרים מחושב כך : $C_w^n = \frac{n!}{(n-w)!\times w!} = \frac{13!}{9!\times 4!} = 715$

10 Mueller, לעיל ה"ש 4.

11 זאת, על פי המתואר אצל Myers, לעיל ה"ש 3.

12 Mueller, לעיל ה"ש 4, בעמ' 103.

13 Bruce S. Weir, *The Rarity of DNA Profiles*, 1 Ann. Appl. Stat 358 (2007).

14 כאשר חזרנו על חישוביו, במודל שלנו, קיבלנו תוצאות זהות.

15 Weir, לעיל ה"ש 13, בטבלה 2 בעמ' 364.

שלו, הפרופילים באריזונה לא רק תואמים במספר אתרים גדול אלא שצפוי שהם יהיו תואמים גם באלל בודד מתוך שניים באותם אתרים נוספים המשלימים ל-13, עד לזהות ב-23 אללים מתוך 26. מצד שני התחזית שלו מנבאת 538 התאמות ב-9 אתרים מתוך 13, מספר הגבוה משמעותית ממה שנמצא בפועל – 122. מתוך אותם 538 התאמות, 50 תואמים חלקית בכל 4 האתרים הנוספים, 163 תואמים חלקית ב-3 מתוך 4 האתרים הנוספים, 198 תואמים חלקית ב-2 מתוך 4 האתרים הנוספים, 106 תואמים חלקית באתר אחד מתוך 4 האתרים ו-21 פרופילים שלא תואמים חלקית באף אחד מארבע האתרים האחרים. מהתוצאות של Weir עולה, כי כאשר פרופיל תואם לפרופיל אחר במספר חלקי גדול של אתרים, רוב הסיכויים שהוא יתאים חלקית באחד או יותר מבין האתרים בהם אין התאמה מלאה.

גם Myers השתמש במודל תיאורטי כדי לחקור את התוצאות שהתקבלו במאגר באריזונה, אלא שבנוסף לחישוב של Weir הוא התייחס גם למספר אוכלוסיות ייחוס.[16] התוצאות שלו, עבור $\theta = 0.00$ מנבאות 100 התאמות בתשעה אתרים ו-3 התאמות בשלושה אתרים (כמו Mueller). עבור ערך $\theta = 0.03$, מתקבלות 13 התאמות ב-10 אתרים ו-320 התאמות ב-9 אתרים. המסקנה של Myers היא שהמצאות קרובי משפחה בדרגת אחים מלאים במאגר, מסבירה את הפער בין התחזית לבין הערך שנמצא בפועל. מסקנתו של Myers בוקרה על ידי Mueller מאחר והחישוב התיאורטי עבור תיקון טטה, מעצם טיבעו, מניח כי הסתברות התאמה עבור פריטים בתוך אותה אוכלוסייה, יהיו רק בין כאלה השייכים לאותה תת-אוכלוסייה בעוד שייתכנו פרופילים תואמים שאינם משתייכים לאותה תת-אוכלוסייה, ובאלה הסתברות התאמה נמוכה יותר.[17] לדעתו, זאת הסיבה למספר הרב של התאמות חלקיות המתקבלות מהמודל התיאורטי של תיקון טטה בתשעה אתרים, יחסית למספר נמוך בהרבה של התאמות בפועל. בסימולציה אותה ביצע Mueller, הוגרלו פרופילים במספר תת-אוכלוסיות ועל כן ההשוואה נעשתה בתוך אותן תת-אוכלוסיות וגם בין תת-אוכלוסיות שונות.

במאמר זה אנו מיישמים מודל תיאורטי הדומה לזה של Myers ו-Weir, לצורך חיזוי תוצאות לגבי המאגר והאוכלוסייה בישראל. בחרנו בגישה התיאורטית כיוון שתיקון טטה הוא המשמש את מומחי התביעה בישראל לצורך עדות בבית המשפט. בכך, אפשר לבחון אותו ואת תקפותו אל מול תוצאות התאמות חלקיות שיכולות להתקבל ביחס למאגר בישראל.

ג. המודל התיאורטי עבור התאמות חלקיות בין פרופילים גנטיים המופקים בעשרה אתרים בעזרת הערכה SGM PLUS בישראל ומגבלותיו

במחקר קודם, תואר מודל תיאורטי לניבוי מספר ההתאמות המלאות בתוך מאגר הכולל אוכלוסייה אחת.[18] התאמות מלאות משמעותן היא שהפרופילים הגנטיים המושווים זהים בכל האתרים הנבדקים. כבמחקר הקודם, ההנחה היא אין שגיאות מעבדה ואין התייחסות לקרבת משפחה. אולם אנו מרחיבים מודל זה עצמו למאגר המורכב ממספר אוכלוסיות, ומחשבים את המספר הצפוי של התאמות חלקיות. כמו כן מובא בחשבון תיקון θ לתת-אוכלוסיות. הפרטים המתמטיים של המודל מוצגים למטה, בנספח הטכני למאמר. ההנחות עליהן המודל מתבסס מפורטות, וצוינו מגבלות מסוימות ביישומו לתחזית התאמות בישראל.

כשבוחנים התאמות מלאות או חלקיות בין פרופילים במאגר המכיל מספר אוכלוסיות יש להבחין בין שני סוגי התאמות. הסוג הראשון הוא כאשר פרופיל גנטי של תורם מאוכלוסייה מסוימת (למשל היהודית) מתאים באופן מלא או חלקי לפרופיל של תורם אחר מאותה אוכלוסייה. התאמות אלה מכונות כאן בשם "התאמות פנימיות". הסוג השני הוא "התאמות חיצוניות", כאשר פרופיל של תורם מאוכלוסייה אחת תואם באופן מלא או חלקי לפרופיל של תורם מאוכלוסיית ייחוס אחרת. מספר ההתאמות הכולל הוא הסכום של שני סוגי התאמות אלה. אבחנה זאת מאפשרת לחשב את מספר ההתאמות במאגר פרופילים המורכב מתורמים ממספר אוכלוסיות. כאשר מתייחסים להתאמות חלקיות יש כמובן להביא בחשבון את העובדה שיש אפשרויות רבות לבחירת מספר חלקי של אתרים מתוך כלל האתרים הנבדק.

16 Myers, לעיל ה"ש 3.
17 Mueller, לעיל ה"ש 4, בעמ' 105.
18 הלפרט, לעיל ה"ש 7.

בנוסף, מקובל לדבר גם על הסתברות הזהות (Probability of Identity).[19] משמעות המושג היא ההסתברות **ששני אנשים אקראיים** מתוך האוכלוסייה יהיו בעלי פרופיל גנטי אחיד.[20] מדובר למעשה על הסתברות התאמה מקרית ממוצעת, ולא על הסתברות התאמה מקרית של פרופיל מסוים הנמצא בידינו לפרופילים אחרים באוכלוסייה. הסתברות זאת רלוונטית למספר ההתאמות הפנימיות הצפוי באוכלוסייה.[21] הסתברות זאת עבור פרופילים בני עשרה אתרים המופקים בערכה SGM PLUS לפריטים מהאוכלוסייה הלבנה בארה"ב (Caucasian), כפי שחושבה על ידי היצרן, היא 2.99×10^{-13}.[22] הסתברות זאת עבור אוכלוסיית African-American היא 7.91×10^{-14}.[23] אוכלוסיות שונות יכולות להיבדל בערך הסתברות הזהות. כך למשל, החישובים שלנו מראים כי הסתברות הזהות עבור 13 אתרים המשמשים להפקת פרופילים במאגר בארה"ב (CODIS), עבור Native Americans משבט ה-Apache שבארה"ב ועבור $\theta = 0.00$ היא 7.6×10^{-13}.[24] זאת בהשוואה להסתברות הזהות עבור האוכלוסייה ה-Caucasian שבארה"ב אשר ערכה, כפי שחושב על ידינו היא 1.76×10^{-15}.[25] החישובים שלנו מראים כי אם אותם 65,493 תורמים במאגר באריזונה היו כולם משבט ה-Apache, ועבור $\theta = 0.00$, אזי היו צפויים שם 5,884 התאמות ב-9 אתרים מתוך 13 (בהשוואה לאותם 122 התאמות שנמצאו בפועל במאגר, ולאותם 100 התאמות שהסימולציה והמודל התיאורטי צופה) ו-294 התאמות ב-10 אתרים (בהשוואה ל-20 התאמות שנמצא בפועל ול-3 התאמות שהסימולציה והמודל התיאורטי צופה). מכאן שמידת הדמיון הגנטי הפנימי באוכלוסייה משתנה כתלות באוכלוסייה.

כאשר מדובר על תת-אוכלוסיות[26], היינו קבוצה של פריטים המתבדלת ברבייה מאוכלוסיית הייחוס, ובפרט כזאת שנהוגים בה נישואיי קרובים, נצפה למידת דמיון פנימית גבוהה משמעותית מזאת הצפויה באוכלוסייה הכללית. זאת, מאחר ומאגר האללים הכללי בתת-האוכלוסייה מצומצם יחסית לאוכלוסיית הייחוס. היינו, הסתברות הזהות בתת-האוכלוסייה תהיה גבוהה מההסתברות הזהות באוכלוסייה הכללית ותוביל בהכרח למידת ההתאמות פנימית גבוהה מאשר בקרב אוכלוסיית הייחוס.

כאן המקום להזכיר את חוק Hardy–Weinberg (להלן "HW"), חוק בסיסי בגנטיקה של אוכלוסיות המגדיר מצב של שיווי משקל הרדי ויינברג כאשר שכיחות הגנוטיפים (צירופי האללים באתר נתון) באוכלוסייה, נגזרת משכיחות האללים באוכלוסייה. על פי חוק זה אוכלוסייה מגיעה לשיווי משקל לאחר דור אחד של זיווגים באקראי. בתנאי חוק שיווי המשקל של HW, האללים בתוך אתר אינם תלויים אחד בשני.[27] חוק זה מתקיים בהנחות הבאות: האוכלוסייה אליה מתייחסים היא אינסופית, הרבייה אקראית (זיווגים באקראי) ואין כוחות מפריעים כסלקציה ברבייה (העדפה לזיווג בין דומים ו/או בין שונים), הגירה ומוטציות.[28] מחוק זה עולה כי כאשר דוגמים אוכלוסייה מסוימת באתר מסוים, אין צורך לדגום את שכיחות כל הקומבינציות של זוגות האללים האפשריים באתר (עניין המצריך מדגם גדול יחסית) אלא ניתן להסתפק בדגימת אללים בודדים ועל סמך שכיחותם, לחזות את שכיחות הצירופים של זוגות האללים באתר.[29] כאשר מוסיפים להנחות אלה גם את ההנחה שחלפו אין סוף דורות מדור המייסדים, מקבלים Linkage Equilibrium - אי תלות בין אללים באתרים שונים, דבר המשמש כהצדקה לשימוש בכלל המכפלה עבור חישוב הסתברות התאמה מקרית בכמה אתרים.[30]

19 Applied Biosystems, AmpFl STR® SGM Plus® PCR Amplification Kit User Manual, 14.12 (2006). available at: http://www3.appliedbiosystems.com/sup/URLRedirect/index.htm?xDoD=4309589

20 שם.

21 ראו משוואות (6) ו-(10) בנספח הטכני למאמר זה.

22 Kit User Manual לעיל ה"ש 19 בעמ' 12-14.

23 שם.

24 Bruce Budowle at all, *CODIS STR Loci Data from 41 sample populations*. 46 J. Forensic Sci. 453 (2001). טבלאות האללים נלקחו מ:

25 Bruce Budowle at al., *Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, U.S. Caucasians, Hispanics, Bahamians*, 44(6) J Forensic Sci. 1277 (1999) טבלאות האללים נלקחו מ:

26 John Buckleton at all, Forensic DNA Evidence Interpretation 3.2.3.2 Crc Press (2005)

27 Forensic DNA Evidence Interpretation, לעיל ה"ש 26, בס' 3.2.1.

28 שם.

29 הביטוי המתמטי לחוק HW, אשר מחשב את שכיחות זוג אללים מסוים באתר מסוים מתוך טבלת שכיחות אללים בודדים באותו אתר נתון במשוואה (2) בנספח הטכני.

30 Forensic DNA Evidence Interpretation, לעיל ה"ש 26, בס' 3.2.2.

טבלאות האללים בישראל מתייחסות רק לאוכלוסייה היהודית והערבית כאוכלוסיות ייחוס. אין לכך לדעתנו הצדקה שכן סביר שבקרב שתי אוכלוסיות אלה מצויות תת-אוכלוסיות הנבדלות אחת מן השנייה. כך למשל בקרב אוכלוסיית היהודים בישראל יש רבים ילידי יבשות שונות, שהיו מנותקים אלו מאלו במהלך ההיסטוריה. תופעות כאפקט מייסדים[31] (founder effect) וסחיפה גנטית (genetic drift), יוצרות שוני גנטי בין האוכלוסיות השונות. אולם כאשר מגדירים כאוכלוסייה אחת צירוף של שתי אוכלוסיות נפרדות, אשר כל אחת מהן בנפרד מקיימת את תנאי שיווי המשקל שבחוק HW, אין הכרח שבאוכלוסייה המצורפת יתקיים, שיווי המשקל של HW.[32] מכאן שדגימת תת-אוכלוסיות שונות בקרב אוכלוסיית ייחוס מרכזית אחת (למשל האוכלוסייה הערבית) והתייחסות אליהן כאוכלוסייה אחת תוביל לסטייה מחוק HW. בנוסף, אין בטחון כי הדמיון הגנטי בקרב תת-אוכלוסייה, כפי שהוא בא לידי ביטוי בהסתברות הזהות, אינו גבוה משמעותית מאשר באוכלוסייה הכללית. לכן, דגימת האוכלוסייה היהודית והערבית בלבד ללא התייחסות לתת-אוכלוסיות מרכזיות בהן, היא הנחת עבודה שגויה.

העולם הפורנזי מנסה להתמודד עם בעיה זאת של תת אוכלוסיות בעזרת שימוש בתיקון θ (טטה).[33] בתיקון זה, מניחים כי בתת-אוכלוסיות קיים קשר בין האללים באותו אתר ואין זה נכון כי הם אינם תלויים. התיקון מעניק הסתברות גבוהה יותר לאתרים הומוזיגוטיים (כאלה ששני האללים בהם זהים) בתת-האוכלוסייה, יחסית להסתברותם באוכלוסיית הייחוס. זאת על חשבון הסתברות האתרים ההטרוזיגוטיים (כאלה שהאללים בהם שונים) בתת-האוכלוסייה.[34] ככל שערך טטה גבוה יותר כך התיקון משמעותי יותר. עבור $\theta = 0.00$ התיקון אינו מתבצע והשימוש בו מתלכד עם חוק HW. בגישה זאת חישוב הסתברות התאמה מקרית בין פרופיל של חשוד לפרופיל השייך לאדם אחר באותה תת-אוכלוסייה אליה משתייך החשוד, וחישוב הסתברות הזהות בתוך אוכלוסיית יחוס מתבצע תחת ההנחה כי עובדת המצאות אלל מסוים בתת-האוכלוסייה, מגדילה את ההסתברות להמצאות אלל נוסף.[35] כלומר אם בתת-אוכלוסייה מסוימת יש תלות בין האללים המצויים באוכלוסייה באותו האתר הרי המצאות אלל מסוים בפרופיל של אדם מגדילה את ההסתברות של המצאות אלל נוסף כזה באותו אדם או באדם אחר באוכלוסייה.[36] מכאן שהתיקון הנהוג אינו קשור לתת-אוכלוסייה מסוימת. לצורך התיקון לא נדרש לדעת את שכיחות האללים בקרב אותה תת-אוכלוסייה. התיקון רק מגדיל את ההסתברות לקבל התאמה בתת-האוכלוסייה אליה משתייך החשוד.

מודל תיקון טטה מטבע הדברים זכה גם לביקורות ולדיונים בספרות.[37] השאלה הפתוחה היא האם תיקון טטה אכן תקף מדעית, והאם הוא אכן עונה על הבעיות המתעוררות כתוצאה מקיום תת-אוכלוסיות. ממחקרו של Weir עולה התחזית התיאורטית עבור טטה בשיעור של 3% היא ל-20 התאמות ב-10 אתרים מתוך 13 כפי שנצפה בפועל.[38] אולם מספר ההתאמות הצפוי ב-9 אתרים מתוך 13 הוא מעל ל-500, הרבה יותר ממה שנמצא בפועל. כאמור במאמרו של Mueller, בטווח ערכים מסוים של ערכי טטה והימצאות מספר מסוים של פרופילים גנטיים של אחים מלאים, ניתן למצוא הסבר להתאמות בתשעה ובעשרה אתרים. אלא שמדובר על פרמטרים שהותאמו במיוחד למאגר באריזונה עבור התאמות בתשעה ועשרה אתרים. כדי לבחון האם באמת התיאוריה הגנטית המקובלת מסבירה את התוצאות באריזונה, יש לדעת נתונים אמיתיים על התורמים למאגר. גם ספירת מספר ההתאמות החלקיות במספר קטן יותר של אתרים, למשל בשמונה ובשבעה אתרים יכול לסייע בקביעה האם אותו שילוב מסוים של פרמטרים מנבא נכון את התוצאות המתקבלות במאגר.

31 Alan R. Templeton, *The theory of speciation via the founder theory*. 94 Genetics 1011 (1979)
32 Forensic DNA Evidence Interpretation, לעיל ה"ש 26 בס' 3.2.3.5.
33 שם, בס' 3.3.3.
34 ראה משוואה (29) בנספח הטכני.
35 ראה משוואה (30) בנספח הטכני.
36 שם, בס' 3.3.3.

> **"However, we invariably have the information that at least one copy of the profile exists. We have seen it in the suspect. In other words, we are not talking about the vast majority of profiles that do not exist; we are talking about one of the few that do, indeed, exist in the real world."… "The main reason for this is population subdivision and relatedness. <u>The fact that one person has the profile slightly increases the probability that his/her relatives members of his/her subpopulation have the profile.</u>"**

ראה גם הנספח טכני
37 Forensic DNA Evidence Interpretation, לעיל ה"ש 26, בס' 3.4.
38 Weir, לעיל ה"ש 13, והטקסט מעל.

במאמר זה, אנו משתמשים בתיקון טטה לצורך חיזוי מספר ההתאמות החלקיות הצפוי במאגר ובאוכלוסייה. המודל שלנו, כאשר יישמנו אותו למאגר באריזונה, תואם בתוצאותיו לתוצאותיהם של Weir ו-Myers. מכאן, שהוא מתאים למודל הגנטי המקובל בקהילה הפורנזית. ביצוע ההשוואה במאגר על ידי הרשויות בישראל וספירת ההתאמות החלקיות במספר רב של אתרים יכול להוות מבחן הפרכה לאותו מודל המתבסס על אותו תיקון טטה. חשוב לזכור כי קיום של תת-אוכלוסיות צפוי להגדיל את מספר ההתאמות הפנימיות הצפויה בתוך אותה תת-אוכלוסייה, יחסית למספר ההתאמות באוכלוסייה הכללית. אולם הדבר לא בהכרח יגדיל את מספר ההתאמות באוכלוסייה כולה. הסיבה לכך היא שסביר שמספר ההתאמות החיצוניות בין תת-האוכלוסיות השונות באופן גנטי, יפחת. זאת מאחר ופריטים מאוכלוסיות שונות יהיו דומים פחות מאשר פריטים מאותה אוכלוסייה.[39]

עניין אחר קשור לטבלאות שכיחות האללים באוכלוסייה. הטבלאות ששימשו אותנו במחקר זה הן טבלאות שהתביעה מעבירה להגנה בתיקי עבודה בבתי משפט בישראל. בטבלאות אלה נראה כי השכיחות עבור אללים נדירים נחסמה מלמטה על ידי הערך המספרי 0.01. דהיינו, אם שכיחות אלל מסוים באתר מסוים נמצאה במדגם המסוים כנמוכה מ-1%, היא נרשמה כ-1%. הדבר מוצדק בחישובי הסתברות התאמה מקרית בתיקי עבודה, מאחר וכאשר המדגם קטן עד למספר מאות פרטים בלבד, הסטטיסטיקה בדבר אללים ששכיחותם נדירה אינה תקפה. מצד שני, הדבר מוביל לכך שטבלאות שכיחות האללים אינן מנורמלות. דהיינו סכום הסתברויות שכיחות האללים באתר מסוים גדול מ-1. אנו סבורים כי יש להגדיל את המדגם כך שגם שכיחות האללים הנדירים תהיה תקפה. לצורך מאמר זה נְרַמֶלנו את טבלאות שכיחות האללים. אולם כנראה שבכך הכנסנו טעות לחישוב, שכן שכיחות האללים הנפוצים קטנה על חשבון ההגדלה המלאכותית של שכיחות האללים הנדירים. למעשה מורידה פעולה זאת את הסתברות הזהות.

מאחר ואין בידינו נתונים ביחס לאוכלוסיות התורמות למאגר הגנטי בישראל, השתמשנו בתחזיות בנתונים מהלשכה המרכזית לסטטיסטיקה. לצערנו, לא מצאנו נתונים לגבי ההרכב הדמוגרפי של אוכלוסיית בתי הסוהר. לכן השתמשנו בסטטיסטיקה של הרשעות בפלילים של מבוגרים, לפי תכונות נבחרות, לשנת 2006.[40] הנחנו כי מאנשים שהורשעו בגין עבירות קלות, כעברות תעבורה והנכללים גם הם בסטטיסטיקה, לא נלקחו דגימות ועל כן אין הם מופיעים במאגר. לכן, בחרנו מהטבלאות רק את אלה שהורשעו ונגזר עליהם עונש מאסר או מאסר וקנס. מאחר וברשותנו טבלאות שכיחות אללים רק לגבי האוכלוסייה היהודית והערבית, בחרנו מהטבלה נתונים רק על אוכלוסיות אלה. למיטב ידיעתנו, שכיחות האללים באוכלוסייה "האחרת" (שאינה ערבית או יהודית) החיה בישראל לא נדגמה עדיין. לכן, השמטנו את אלה המהווים אוכלוסייה אחרת החיה בישראל. הפרופורציות במאגר חושבו בטבלה הבאה:

טבלה 1: פרופורציות משוערות של הרכב מאגר גנטי בישראל

	אוכלוסייה יהודית	אוכלוסייה ערבית	סה"כ
1. מספר אלו שהוטלו עליהם עונשי מאסר	5,635	2,811	8,446
2. מספר אלה שהוטלו עליהם עונשי מאסר פלוס קנס כספי.	1,762	1,362	3,124
סיכום 1 עם 2	7,397	4,173	11,570
פרופורציות	64%	36%	100%

מאחר ומטרת ההשוואה למאגר היא רק בדיקת התיאוריה הגנטית, הפרדנו בין גברים לנשים. אם תתבקש השוואת תוצאות התחזית שלנו לספירת ההתאמות במאגר, יהיה צורך להפחית את אלה שאינם שייכים לאוכלוסייה היהודית והערבית. אנו נשמח לקבל נתונים אמיתיים על המאגר באופן שנוכל לתקן את התחזיות שלנו בעזרת נתונים אמיתיים על הרכב התורמים למאגר.

39 ראו ה"ש 17 והטקסט מעל.

40 ראו הרשעות במשפטים פליליים לפי תכונות נבחרות אצל הלשכה המרכזית לסטטיסטיקה, שנתון סטטיסטי לישראל, 491 (2008). העמוד המסוים ניתן לצפייה כאן: http://www.cbs.gov.il/shnaton59/st11_04.pdf

את האוכלוסייה בישראל הערכנו על פי נתוני הלשכה המרכזית לסטטיסטיקה שבשנתון הסטטיסטי לישראל לשנת 2008.[41] לוח 2.11 בשנתון מביא מידע על אוכלוסיית ישראל בשנת 2007 על פי חתכים של דת, גיל ומין.[42] תחזית ההתאמות היא רק עבור האוכלוסיות היהודית והערבית. בהנחה כי הפרופיל הגנטי המופק מזירת הפשע מאפשר לדעת את מין מבצע העבירה, הפרדנו בין ההתאמות הצפויות באוכלוסיית הגברים לבין אלה הצפויות באוכלוסיית הנשים. התמקדנו באוכלוסייה מעל גיל 15, בהנחה (ותקווה) כי צעירים יותר אינם קשורים לפשיעה.

טבלה 2: פרופורציות של האוכלוסייה היהודית והערבית בישראל מעל גיל 15 לפי מין בשנת 2007

אוכלוסייה	גברים (באלפים)	נשים (באלפים)
יהודים	1956.2	2077.9
ערבים	435.4	425.6
סה"כ	2391.6	2503.5
פרופורציה ערבים	17.2%	17.0%
פרופורציה יהודים	82.8	83.0%

מספר ההתאמות בפועל בכלל האוכלוסייה בישראל צפוי להיות גבוה במקצת מהתחזית שלנו, שכן אמורות להיכלל בה גם התאמות פנימיות בין אוכלוסיית ה"אחרים" לבין עצמם, והתאמות חיצוניות בין אוכלוסיית ה"אחרים" לאוכלוסיות היהודים והערבים. שיעור אוכלוסייה זאת בכלל אוכלוסיית ישראל הוא קטן, אולם כפי שעולה מהנתונים הסטטיסטיים, מעורבותם של פרטים מאוכלוסייה זאת בפשיעה, גבוה יותר משיעורם באוכלוסייה.[43]

נציין גם כי בחישוב התיאורטי אנו סופרים את מספר הזוגות בלבד. ייתכן כי יהיו מספר בני אדם ולא רק זוג, להם אותו פרופיל גנטי חלקי. בשיטת הספירה שלנו, שלישיית פרופילים חלקיים הזהה במספר חלקי של אתרים, תיחשב כשלושה זוגות פרופילים תואמים חלקית. רביעייה כזאת תחשב כ-6 זוגות פרופילים תואמים חלקית וכו'.[44]

ד. השוואה בין התוצאות מהמודל התיאורטי לתוצאות סימולציה ממוחשבת

מטרת חלק זה היא לבדוק את המודל התיאורטי, אל מול זה המתקבל מסימולציה ממוחשבת. בסימולציה, פרופילים גנטיים מהאוכלוסייה היהודית ומהאוכלוסייה הערבית הוגרלו בעזרת מחולל מספרים אקראיים על פי טבלאות שכיחויות אללים מתאימות.[45] הסימולציה נעשתה בשפת התכנות C++.

כאמור, אם N מסמל את המספר הכללי של הפרופילים במאגר, נדרשות $\frac{N\times(N-1)}{2}$ פעולות השוואה. עבור N גדול, ההשוואות עשויות להימשך זמן רב. במאמר אחר, תואר שימוש ב-Hash Table להאצת החישוב.[46] בשיטה זאת, פרופילים הזהים במספר אתרים מסוים ומוגבל אוגדו בקבוצות חלקיות כך שפרופיל הנמצא בקבוצה אחת אינו יכול להתאים לפרופיל מקבוצה אחרת. לכן, מספיק להשוות בין פרופילים מאותה קבוצה, פעולה המאיצה את החישוב. אולם שיטה זאת אינה ישימה עבור התאמות חלקיות שכן התאמות חלקיות אפשריות גם בין הקבוצות השונות. לכן, ממגבלות זמן הרצה, בחרנו לבצע את הסימולציה על מאגר קטן המונה 1000 פריטים ועבור $\theta = 0.00$. ביצענו את הסימולציה 1000 פעם כדי לקבל סטטיסטיקה טובה והצבנו את הערכים הממוצעים. השווינו את התוצאות מהמודל התיאורטי לאלה מהסימולציה עבור שלושה מקרים שונים: א. בהנחה שכל N=1000 הפרופילים הם מהאוכלוסייה

41 הלשכה המרכזית לסטטיסטיקה, שנתון סטטיסטי לישראל (2008). ניתן לצפייה באתר:
http://www.cbs.gov.il/shnaton59/shnaton59_all.pdf

42 השנתון שלעיל, בעמ' 122-116, לוח 2.11

43 ראו הרשעות במשפטים פליליים לפי תכונות נבחרות, לעיל ה"ש 40.

44 קבוצה של n פרופילים גנטיים זהים כוללת בתוכה $z=\frac{n!}{(n-2)!\times 2!}$ זוגות זהים.

45 מחולל המספרים האקראיים בו השתמשנו הוא גרסה בשפת התכנות c למחולל מספרים אקראיים המתואר אצל:
George Marsaglia and Arif Zaman, *Toward a Universal Random Number Generator*, Florida State University Report: FSU-SCRI-87-50 (1987) available at: www.jud10.org/AdministrativeOrders/LocalRules/RandomNumberGenerator.pdf
גרסת שפת c, לאלגוריתמים זה, זמינה כאן: http://local.wasp.uwa.edu.au:80/~pbourke/other/random/randomlib.c

46 הלפרט, לעיל ה"ש 7, בעמ' 115.

היהודית. ב. בהנחה שכל N=1000 הפרופילים הם מהאוכלוסייה הערבית. ג. בהנחה שהמאגר מורכב מתורמים מהאוכלוסייה היהודית והערבית על פי הפרופורציות הנתונות בטבלה 1.

טבלה 3: השוואה בין המודל התיאורטי לבין הסימולציה עבור מספר התאמות חלקיות הצפוי בין פרופילים גנטיים שהופקו ב-10 אתרים במאגר בו N=1000 פרופילים גנטיים

מספר אתרים בהם צפויה התאמה	מספר התאמות במודל **התיאורטי** באוכלוסיה היהודית	מספר התאמות **בסימולציה** באוכלוסיה היהודית	מספר התאמות במודל **התיאורטי** באוכלוסיה הערבית	מספר התאמות **בסימולציה** באוכלוסיה הערבית	מספר התאמות במודל **התיאורטי** באוכלוסיה מעורבת (64% 36%)	מספר התאמות **בסימולציה** באוכלוסיה מעורבת (64% 36%)
6	1.189	1.165	0.841	0.895	0.985	1.001
5	28.841	28.928	23.305	23.312	25.548	25.499
4	484.383	485.192	409.071	410.305	438.053	439.422
3	5428.83	5433.063	4815.707	4822.393	5057.364	5059.034

הטבלה מראה כי ככל שמספר ההתאמות הצפוי גדול יותר כך מתקבלת התאמה טובה יותר בין המודל התיאורטי לסימולציה.[47] בכך וידאנו כי המודל התיאורטי מתאר נכון את המודל הגנטי עבור התאמות חלקיות בין מספר אוכלוסיות ייחוס.

ה. תחזית להתאמות חלקיות במאגר פרופילים בישראל כפונקציה של גודל המאגר

לאחר שווידאנו כי התיאוריה מתאימה לסימולציה נוכל להשתמש במודל התיאורטי, במגבלות ההנחות עליהן הוא מבוסס, כדי לחזות את מספר ההתאמות הצפוי במאגר גנטי. הסתברות הזהות (Probability of Identity) עבור האוכלוסייה היהודית בישראל על פי הטבלאות המצויות בידינו לעֶרכַּת SGM PLUS ו- $\theta = 0.00$, היא 4.61×10^{-14}. הסתברות הזהות עבור האוכלוסייה הערבית בישראל היא 2.94×10^{-14} . עבור $\theta = 0.03$ הסתברות הזהות באוכלוסייה היהודית היא 6.89×10^{-13} ועבור האוכלוסייה הערבית היא 4.78×10^{-13} .
אלה הסתברויות התאמה מאד נמוכות, אשר לכאורה חוזות יכולת הפרדה גבוהה.[48] להשוואה, הסתברות הזהות עבור האוכלוסייה ההיספאנית בארה״ב, כפי שחושבה על ידינו עבור 13 אתרים הנבדקים בארה״ב, עבור $\theta = 0.00$ היא 4.52×10^{-15}.[49] כאמור, עבור אמריקאים ״ילידים״ משבט ה-Apache הסתברות הזהות היא 7.6×10^{-13}.[50]

מאחר ואין אנו יודעים את גודל המאגר בישראל, ומטבע הדברים הוא גדל עם השנים, אנו חוזים התאמות לגדלים שונים של המאגר. ביצענו זאת בהנחה כי המאגר מכיל פרופילים מיהודים וערבים לפי הפרופורציות הנתונות בטבלה 1 (64% יהודים ו-36% ערבים) ועבור ערכי טטה שונים.

47 הדבר הקובע בפועל את מספר ההתאמות במאגר הוא Probability of Identity או המקבילה שלו בחישוב התאמות החלקיות (ראו נספח טכני). בחנו גם את ערכה הממוצע של הסתברות זאת מאותם אלף ניסויים למול הערך התיאורטי, וביחס לסטיית התקן באותם סימולציות. למשל, הערך התיאורטי של הסתברות זאת כפי שחישבנו עבור האוכלוסייה היהודית הוא 4.61×10^{-14}. הערך הממוצע שהתקבל בסימולציה עבור 640 פרופילים מתורמים יהודיים היה 4.56×10^{-14} כאשר סטיית התקן באותם 1000 ניסויים הייתה 1.12×10^{-14} מבלי להיכנס למבחנים סטטיסטים או לחישוב תיאורטי של השגיאה שעדיין לא פורסמו בספרות ואשר ראויים למאמר בפני עצמו, זאת בהחלט התאמה טובה.

48 כאמור, הנרמול שנאלצנו לעשות לטבלאות המצויות בשימוש בידי מומחי התביעה בישראל, מכניס גורם שגיאה ומוריד את הסתברות הזהות.

49 טבלאות האללים נלקחו מ-Bruce Budowle at all, לעיל ה״ש 25.

50 טבלאות האללים נלקחו מ-Bruce Budowle at all, לעיל ה״ש 24.

טבלה 4: מספר ההתאמות החלקיות בין פרופילים גנטיים שהופקו ב-10 אתרים הצפוי במאגר הכולל N פרופילים מתורמים יהודיים וערבים עבור $\theta = 0.00$

מספר אתרים בהם צפויה התאמה	מספר התאמות צפוי עבור N=40000	מספר התאמות צפוי עבור N=60000	מספר התאמות צפוי עבור N=80000	מספר התאמות צפוי עבור N=100000	מספר התאמות צפוי עבור N=400000
10	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.7
8	0.7	1.6	2.8	4.4	70.3
7	42.0	94.5	167.9	262.4	4198.1
6	1602.6	3605.9	6410.6	10016.5	160265.5
5	40841.0	91893.1	163366.2	255260.4	4084199.0

טבלה 5: מספר ההתאמות החלקיות בין פרופילים גנטיים שהופקו ב-10 אתרים הצפוי במאגר הכולל N פרופילים מתורמים יהודיים וערבים עבור $\theta = 0.015$

מספר אתרים בהם צפויה התאמה	מספר התאמות צפוי עבור N=40000	מספר התאמות צפוי עבור N=60000	מספר התאמות צפוי עבור N=80000	מספר התאמות צפוי עבור N=100000	מספר התאמות צפוי עבור N=400000
10	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.1	0.1	1.7
8	1.5	3.4	6.1	9.6	153.2
7	79.7	179.4	318.9	498.3	7973.0
6	2677.3	6024.0	10709.4	16733.4	267737.7
5	60650.1	136464.2	242604.4	379070.6	6065190.0

טבלה 6: מספר ההתאמות החלקיות בין פרופילים גנטיים שהופקו ב-10 אתרים הצפוי במאגר הכולל N פרופילים מתורמים יהודיים וערבים עבור $\theta = 0.03$

מספר אתרים בהם צפויה התאמה	מספר התאמות צפוי עבור N=40000	מספר התאמות צפוי עבור N=60000	מספר התאמות צפוי עבור N=80000	מספר התאמות צפוי עבור N=100000	מספר התאמות צפוי עבור N=400000
10	0.0	0.0	0.0	0.0	0.0
9	0.0	0.1	0.2	0.3	4.9
8	3.7	8.3	14.8	23.1	369.7
7	164.6	370.3	658.3	1028.6	16458.0
6	4765.1	10721.6	19060.8	29782.6	476527.8
5	93991.7	211484.0	375974.0	587461.6	9399492.0

הטבלאות מראות כי כבר במאגר המונה 40,000 פרופילים גנטיים, ומבלי להתייחס לתיקון טטה, ייתכן שנראה התאמה ב-8 אתרים (טבלה 4 מנבאת 0.7 התאמות). אם תיקון טטה תקף, ואם נבחר עבורו ערך של 1.5% או אף 3%, ההסתברות לכך גדלה עד ל-3.7 התאמות צפויות. אם בתוך המאגר יש זוגות אחים רבים, אזי התוצאות יכולות לגדול אף בעשרות אחוזים.[51] כאמור, בישראל נדגמות רק אוכלוסיות ייחוס מרכזיות (היהודית והערבית) כאשר לתת-אוכלוסיות (כמו ערבים נוצרים, דרוזים וכו') משתמשים בתיקון טטה. תיקון טטה אמור להוות פתרון גם לאפשרות שהזיווגים אינם אקראיים אלא גברים ונשים בוחרים זה את זו על פי תכונות מסוימות כאינטליגנציה, מיקום גיאוגרפי וכו'. תחזיות

51 ראו Myers, לעיל ה"ש 3.

אלה שלנו יכולות להוות מבחן הפרכה לתקפות תיקון טטה בישראל. לצורך בחינת התחזיות, נדרש מהרשויות לכתוב תוכנת מחשב פשוטה אשר תשווה בין כל הפרופילים הגנטיים במאגר ותספור את מספר ההתאמות החלקיות. יש כמובן להתייחס בספירה רק לפרופילים שהתורמים להם הם מהאוכלוסיות היהודית והערבית בלבד. במספר אתרים גדול, בו צפויות להתקבל במאגר התאמות מעטות יחסית, אנו ממליצים על בדיקה פרטנית של כל התאמה כזאת כדי לאפיין את המקרים. נשמח לקבל מן הרשויות נתונים מדויקים על הרכב אוכלוסיית התורמים למאגר וטבלאות אללים גם לאוכלוסיית האחרות בישראל ולתקן את התחזיות בהתאם.

ו. תחזית עבור מספר התאמות מלאות וחלקיות הצפוי

בפרופילים הגנטיים של האוכלוסייה בישראל

להלן התוצאות עבור מספר ההתאמות המלאות והחלקיות הצפוי באוכלוסייה כולה. התחזית ניתנת רק לאוכלוסייה היהודית והערבית ללא התחשבות בהתאמות מאוכלוסייה אחרת. יש הפרדה בין אוכלוסיית הגברים לנשים וההתמקדות היא בבני 15 ומעלה. על פי טבלה 2, מספר הגברים בישראל מעל גיל 15, כמעודכן לשנת 2007, היה 2.39 מיליון. מתוכם 82.8% יהודים ו-17.2% ערבים. מספר הנשים היה 2.50 מיליון כאשר 83.0% מתוכן יהודיות ו-17.0% ערביות. להשוואה חישבנו עבור $\theta = 0.00$ גם את מספר ההתאמות הצפוי, לו כל האוכלוסייה הייתה יהודית ולו כל האוכלוסייה הייתה ערבית. התחזית ניתנת על סמך ערכי טטה שונים לאוכלוסייה הכללית.

טבלה 7: התאמות חלקיות צפויות בין פרופילים גנטיים בישראל בקרב אוכלוסיית הגברים מעל גיל 15 בישראל
המונה N = 2.3916 מיליון תושבים, עבור $\theta = 0.00$

מספר אתרים בהם צפויה התאמה	מספר התאמות צפוי עבור N כאשר האוכלוסייה כולה יהודית	מספר התאמות צפוי עבור N כאשר האוכלוסייה כולה ערבית	מספר התאמות צפוי עבור N כאשר 82.8% יהודים ו-17.2% ערבים
10	0.1	0.1	0.1
9	30.4	20.3	27.1
8	3075.3	2151.2	2773.6
7	179284.3	131396.5	163649.5
6	6672357.5	5126201.5	6167521.0
5	165627648.0	133470336.0	155122544.0

טבלה 8: התאמות חלקיות צפויות בין פרופילים גנטיים בישראל בקרב אוכלוסיית הנשים מעל גיל 15 בישראל
המונה N = 2.5031 מיליון תושבות, עבור $\theta = 0.00$

מספר אתרים בהם צפויה התאמה	מספר התאמות צפוי עבור N כאשר האוכלוסייה כולה יהודית	מספר התאמות צפוי עבור N כאשר אוכלוסייה כולה ערבית	מספר התאמות צפוי עבור N כאשר 83.0% יהודיות ו-17% ערביות
10	0.1	0.1	0.1
9	33.3	22.3	29.7
8	3368.7	2356.5	3041.7
7	196391.0	143933.9	179444.2
6	7309011.0	5615326.0	6761811.5
5	181431264.0	146205616.0	170044640.0

בטבלה 9 ו-10 מצויה התחזית להתאמות חלקיות בישראל עבור ערכי טטה שונים כאשר פרופורציות האוכלוסיות נלקחו מטבלה 2.

טבלה 9: התאמות חלקיות צפויות בין פרופילים גנטיים בישראל בקרב אוכלוסיית הגברים מעל גיל 15 בישראל
המונה N = 2.3916 מיליון תושבים עבור ערכי θ שונים

מספר אתרים בהם צפויה התאמה	$\theta = 0.01$	$\theta = 0.02$	$\theta = 0.03$
הסתברות הזהות בכל 10 האתרים	אוכלוסייה יהודית 1.20×10^{-13} אוכלוסייה ערבית 7.90×10^{-14}	אוכלוסייה יהודית 2.95×10^{-13} אוכלוסייה ערבית 2.00×10^{-13}	אוכלוסייה יהודית 6.89×10^{-13} אוכלוסייה ערבית 4.78×10^{-13}
10	0.3	0.6	1.4
9	55.3	115.4	238.2
8	5093.8	9552.0	17788.3
7	272126.0	460144.5	775399.5
6	9335347.0	14299947.0	21870530.0
5	214908816.0	299807200.0	417754272.0

טבלה 10: התאמות חלקיות צפויות בין פרופילים גנטיים בישראל בקרב אוכלוסיית הנשים מעל גיל 15 בישראל
המונה N = 2.5031 מיליון תושבים עבור ערכי θ שונים

מספר אתרים בהם צפויה התאמה	$\theta = 0.01$	$\theta = 0.02$	$\theta = 0.03$
הסתברות הזהות בכל 10 האתרים	אוכלוסייה יהודית 1.20×10^{-13} אוכלוסייה ערבית 7.90×10^{-14}	אוכלוסייה יהודית 2.95×10^{-13} אוכלוסייה ערבית 2.00×10^{-13}	אוכלוסייה יהודית 6.89×10^{-13} אוכלוסייה ערבית 4.78×10^{-13}
10	0.3	0.7	1.5
9	60.6	126.5	260.9
8	5579.8	10463.4	19485.6
7	298091.4	504050.0	849385.5
6	10226095.0	15664400.0	23957342.0
5	235414688.0	328413792.0	457614976.0

בטבלאות ניתן לראות שיכולת ההפרדה הלכאורית של הערכה גבוהה. בקירוב, זוג אחד בלבד של פרופילים התואמים בעשרה אתרים צפוי להימצא הן באוכלוסיית הנשים והן באוכלוסיית הגברים בכל ישראל כולה, כאשר איננו מביאים בחשבון קרבת משפחה. נציין כי המודל שלנו, כאשר הוא מיושם עבור המאגר באריזונה, צופה כ-3.4 התאמות בעשרה אתרים בקרב 65,493 פרופילים גנטיים. ההסבר לפער זה נובע לפחות משני גורמים. האחד, במאגר באריזונה מדובר על התאמות של עשרה אתרים מתוך 13. יש 286 אפשרויות לבחור 10 אתרים מתוך 13. בחישוב שלנו, מדובר על השוואת 10 אתרים מתוך 10. שנית, התוצאות לא ניתנות להשוואה שכן מדובר על אתרים שונים שנבדקים, ולהם יכולות הפרדה שונות.

מספר ההתאמות החלקיות הצפוי בתשעה אתרים עבור $\theta = 0.03$ הוא מספר מאות בלבד, הן לאוכלוסיית הנשים והן לאוכלוסיית הגברים. במחקר קודם פורסם כי בישראל צפויים 14,865 זוגות פרופילים שהם זהים באופן מלא ב-9 אתרים.[52] אלא שבמחקר הקודם לא נעשתה אבחנה בין התאמות באוכלוסיית הגברים לאוכלוסיית הנשים. המחקר ההוא התייחס לכלל האוכלוסייה ולא רק לבוגרים בני 15 ומעלה. בנוסף, המחקר ההוא התבסס על אתרים שחלקם שונים מאלה הנבדקים על ידי העֵרכָּה SGM PLUS. הסתברות הזהות באותם 9 אתרים חושבה ל- 6.606750×10^{-10}.[53] אלא שנראה על פניו מחישוב הסתברויות הזהות, מטבלאות שכיחויות האללים שהועברו אלינו, כי האתרים הנבדקים בערכה החדשה הם בעלי יכולת הפרדה גבוהה יותר.[54]

מספר ההתאמות החלקיות הצפוי ב-8 אתרים הוא כעשרים אלף. יש קרוב למיליון פרופילים המתאימים ב-7 אתרים, ומספר גדול מאד של זוגות פרופילים המתאימים ב-6 ויותר אתרים. מתוכם ייתכנו גם פרופילים גנטיים המשותפים למספר בני אדם ולא רק לזוג בני אדם.[55]

למרות שמטבלאות 9 ו-10, עולה כי התיאוריה הגנטית צופה התאמות חלקיות רבות, יש לברר כמובן את המשמעות המשפטית של עובדות אלו.

ז. דיון משפטי כללי

בעניין מוראד אבו חמאד כתב כב' השופט מ' חשין:

> **"שכיחות של אחד בין מיליארדים די בה כדי לקשור פלוני לעבירה וכך אף בשכיחות של מיליונים, ואולם אין צורך שנקבע מסמרות."**[56]

המדובר בפרופיל גנטי שהופק ב-9 אתרים מן העקבות שהשאיר אחריו אנס. האם בעובדה כי באריזונה, לאחד מבין כל 227 בני אדם יש תואם גנטי ב-9 אתרים מתוך 13, כבר בקרב 65,493 איש,[57] יש בה לבדה כדי להטיל ספק במסקנה של השופט חשין? ראשית נאמר כי לדעתנו ההכרעה כאן אינה מדעית כי אם הכרעה בסוגיה של "רלוונטיות". הכרעה בסוגיות של רלוונטיות הן עניין למשפטן ולא לאיש מדע.[58] אולם, כדי שהמשפטן יוכל להכריע בסוגיות אלה, עליו להיות מודע לנתונים המדעיים המהווים בסיס להחלטותיו. להלן טיעונים משפטיים אפשריים, הן מצד התביעה והן מצד ההגנה, המבוססים על אותם נתונים מדעיים.

עניין אחד קשור לעצם הרלוונטיות של תוצאות הנוגעות למספר התאמות הכללי הצפוי באוכלוסייה, אשר צפויות להתקבל כאשר משווים את כל הפרופילים באוכלוסייה בינם לבין עצמם. התביעה יכולה לטעון כי הסתברות ההתאמה המקרית של אחד למיליארדים היא הרלוונטית ולא מספר הזוגות, שלישיות וכו' התואמים באוכלוסייה. זאת מאחר ואין לנו עניין לדעת כמה התאמות יש כאשר אנו משווים את כל הפרופילים הגנטיים של תושבי המדינה לעצמם, אלא שיש לנו עניין לדעת את ההסתברות שהפרופיל המסוים מזירת הפשע, הנמצא בידינו, מתאים לאדם אקראי אחר חוץ מן הנאשם. ההגנה יכולה לטעון כי טענה זאת של התביעה הייתה נכונה לו המשטרה הייתה חוקרת פשע אחד. אולם המשטרה חוקרת פשעים רבים, ובכך היא מחפשת התאמה לפרופילים רבים מזירות פשע רבות, ולא רק עבור פרופיל אחד. ייתכן וצפוי שבאחת מהחקירות הרבות יימצאו באופן מקרי גם שני פרופילים תואמים כפי שאכן מצויים באוכלוסייה.

בעניין זה אנו נוטים להסכים עם עמדת ההגנה. מאחר והמשטרה חוקרת פשעים רבים, יש בכך כדי להחליש את הרושם בדבר ייחודיות הפרופיל הגנטי אשר מחושבת ביחס אליו הסתברות התאמה מקרית של אחד למיליארדים, או אחד

52 הלפרט, לעיל ה"ש 7.
53 הלפרט, לעיל ה"ש 7, בעמ' 100 משוואה (2).
54 ראו דיון בתחילת פרק ה' למאמר זה, על הסתברויות הזהות הנמוכות שחושבו על ידינו עבור האוכלוסייה היהודית והערבית.
55 הלפרט, לעיל ה"ש 7.
56 ע"פ 9724/02 **מוראד אבו-חאמד נ' מדינת ישראל**, דינים עליון כרך ס"ה, 552 (2003) בפס' 35 לפסק-דינו של כב' השופט מ' חשין.
57 לבדה, משמעותה ללא הבאה בחשבון של האפשרות לקבלת פרופיל גנטי שגוי כפי שיידון בפרק ח.
58 ראו קביעה כזאת הנוגעת למחלוקת בדבר משקל ראיית ה-DNA כאשר היא מושגת דרך סריקה במאגר אצל:
United States v. Jenkins, 887 A.2d 1013, 1025 (Ct. App. D.C. 2005)

לטריליונים. מכאן כי ראוי שבית המשפט יתייחס בהחלטותיו גם למספר ההתאמות הצפוי באוכלוסייה כולה. בהקשר זה יש צורך לעשות אבחנה ברורה בין שני מצבים. **א**. כאשר פרופיל מזירת הפשע מושווה לפרופיל אחד של חשוד שאותר על סמך ראיות מסוימות. **ב**. כאשר משווים בין פרופיל מזירת הפשע לפרופיל ממאגר נתונים. הסבירות להתאמה מקרית במצב הראשון נמוכה יותר, מאחר ויש ראיות נוספות הקושרות את הנאשם לעבירה.[59] במצב השני, כאשר אין ראיות נוספות, יש חשיבות לאפשרויות ההתאמה המקרית באוכלוסייה.

אירועים נדירים מבחינה סטטיסטית מתרחשים יום יום. כך למשל ההסתברות שאדם יזכה בהגרלת פיס היא נמוכה. אולם באוכלוסייה גדולה תמיד יהיו כאלה שלמרות הסיכוי הנמוך יזכו בפיס. יתרה מכך, למרות שסטטיסטיקה של מספרים גדולים היא מדויקת, לא כך לגבי התאמה בודדת. התאמה בודדת היא עניין סטוכסטי. גם אם נצפה בממוצע להתאמה מקרית אחת במאגר גדול, כל 1,000 סריקות לדוגמה, זה עדיין רק בממוצע. ייתכן שהתאמה מקרית תתקבל לאחר מספר סריקות, כאשר הסתברות הזהות היא אחת למיליארדים, וייתכן כי היא תתקבל לאחר עשרת אלפים סריקות.[60] דרך אחרת להסביר זאת היא שגם אם בידינו פרופיל גנטי הצפוי להתאים לפרופיל אחד ברשימה של מיליון פרופילים, עדיין ייתכן שההתאמה תימצא כבר עם הפרופיל הראשון ברשימה וייתכן שההתאמה תימצא רק עבור הפרופיל האחרון ברשימה.

עניין נוסף הוא ההתאמות החלקיות. בעניין אבו חמאד, נבדקו רק 9 אתרים, מאלה נמצאה התאמה בכולם. אולם לו היינו בודקים את האוכלוסייה כולה ב-13 אתרים היה מתגלה שרבים באוכלוסייה חולקים התאמות ב-9 אתרים מתוך אותם 13. ההגנה יכולה לטעון כי יש בדבר כדי להחליש את המשקל שיש לתת להתאמה בבדיקת ה-DNA. ההגנה יכולה להיעזר בעניין זה במלומד Tribe שהסתייג, ברמה העקרונית, מכימות הסתברות האשמה.[61] ההגנה גם תוכל לטעון כי מדובר על מחדל בחקירה, שכן לו נבדקו במקרה מוראד אבו חאמד אתרים נוספים לבד מאותם תשעה, ייתכן ולא הייתה נמצאת התאמה בשאר האתרים.[62] התביעה יכולה לטעון כי אכן צפוי שרבים יחלקו 9 אתרים מתוך 13, אולם אלה בדרך כלל אינם אותם 9 אתרים. עובדה היא שמספר ההתאמות הצפוי, כפי שמאמר זה מנבא ב-9 אתרים מתוך 10, נמוך ממספר ההתאמות שנמצאו באריזונה ב-9 אתרים מתוך 13. לעומתה, התביעה יכולה להוסיף ולטעון כי אין היא מונעת מן ההגנה לבדוק את הדגימה באתרים שונים ולכן המחדל הלכאורי בחקירה ניתן ל תיקון.

יש כמובן אמת בטענה אפשרית זאת של התביעה. אולם קשה להתעלם לחלוטין מטענת ההגנה. אנו סבורים כי ניתן לפתור מחלוקת זאת בדרך הבאה - לאחר שמתקבלת החלטה להגיש כתב אישום נגד הנאשם המסוים, ייבדקו אתרים נוספים לא בערכת SGM PLUS אלא בערכה שונה. שילוב של בדיקה בשתי ערכות שונות, יניב בדיקה ב-15 אתרים. עלות בדיקה כזאת למדינה זניחה לעומת הנזק שעלול להיגרם לנאשם שהוא חף מפשע. יתרה מכך, בדרך זאת תיחסך עלות ומאמץ של בדיקת עשרות פרופילים גנטיים נוספים, הנעשית לעתים בתיקי חקירה לצורך אימות ההתאמה לחשוד. מה גם שעלויות בדיקה גנטית לנאשם גבוהות בהרבה מעלויות המדינה. מה גם שנטל ההוכחה הוא על המדינה, ואין היא רשאית להעבירו על הנאשם, שאין עליו נטל הוכחת חפותו.[63] העובדה היא כי בתיקי עבודה רבים, נאשמים אינם מבצעים בדיקה בעצמם אם בגלל שהם אינם יכולים לממן עלות של בדיקה והבאת מומחה מחו"ל להעיד על תוצאות הבדיקה, ואם מסיבות אחרות. לאור הנזק העצום הנגרם מהרשעות שווא, בדיקה באתרים אחרים מהווה אמצעי מניעה שעלותו פחותה מתוחלת הנזק.[64]

האמור בפרק ז' הוא תחת ההנחה הלא נכונה, כי לא תיתכן תקלה בהפקת הפרופיל הגנטי וניתן להתייחס אל הפרופיל כממצא מוחלט. אולם כפי שנראה בסעיף הבא, לא כך הדבר.

---

59 על חשיבות הראיות הנוספות, ראה הלפרט ופרדס, ה"ש 2 לעיל.

60 הלפרט, לעיל ה"ש 7 בעמ' 110.

61 Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329 (1971)

62 ראו מקרה בה נאשם זוכה לאחר שהמשטרה סברה כי מאחר והוא הודה בחקירה, אין צורך לבצע גם מסדר זיהוי. **ע"פ 7443/06** רואי ארקה נ' מדינת ישראל **(אתר בית המשפט העליון 28/9/2008)**

63 Boaz Sangero & Mordechai Halpert, *From a Plane Crash to the Conviction of an Innocent Person: Why Forensic Science Evidence Should Be Inadmissible Unless It Has Been Developed As a Safety-Critical System*, 32 Hamline L. Rev. 65 (2009).

64 שם. ראו גם מרדכי הלפרט ובועז סנג'רו, הסכנה שבהרשעה על סמך נשיפה (על בדיקת הנשיפה המתיימרת לאתר נהיגה בשכרות), **הפרקליט – ספר דיויד ווינר על המשפט הפלילי ואתיקה**, 313, 343 (2009). המאמר ניתן לצפייה בכתובת:
http://www.hapraklit.co.il/_Uploads/dbsAttachedFiles/Sangero1.pdf
http://www.hapraklit.co.il/?CategoryID=219

ח. משמעויות מעשיות – פרופילים חלקיים, פרופילים שגויים ו-allelic dropout בתיקי עבודה.

פרק זה עוסק במשמעות המעשית הנובעת מאפשרות של התאמות חלקיות מקריות בנסיבות שונות כגון הפקת פרופילים חלקיים, או שגויים.

בכל הקשור לפרופילים חלקיים, וכפי שאנו רואים מטבלאות 9 ו-10, יש קרוב למיליון זוגות פרופילים התואמים באוכלוסייה ב-7 אתרים. לכן, למרות הסתברויות התאמה מקרית של מאות מיליונים ואולי אף פחות, באותם שבעה אתרים מהם הופק הפרופיל, אין הדבר אומר כי הפרופיל החלקי הוא ייחודיי. ייתכנו בישראל אף מספר עותקים של אותו פרופיל חלקי המשותף למספר בני אדם.[65]

בעיה נוספת היא האפשרות המעשית להפקת פרופילים שגויים. פרופיל שגוי הוא פרופיל שהופק מדגימה גנטית אשר אינו זהה לפרופיל של התורם לדגימה באלל אחד או יותר. בכל מאגר גדול סביר מאד שיש פרופילים שהופקו באופן שגוי.[66] קשה לגלות פרופיל שגוי שכן אין כלל זהב הקובע מהו הפרופיל האמיתי של התורם לדגימה מסוימת. לכן, הגדרה אחת לפרופיל שגוי באתר היא פרופיל שהופק מדגימה ואשר נצפו בו אללים שונים כאשר הבדיקה התבצעה בשניים או יותר מכשירי מדידה שאינם תלויים זה בזה.[67] שגיאות כאלה התגלו למשל בניו-זילנד, שם באתר TH01 נמצאו האללים 9.3, 10, 9.3. אולם צירוף זה הוא נדיר ועל כן עורר עניין. בדיקה חוזרת הראתה שאלה הם האללים השכיחים יותר 9.3, 10. והמקור למדידה השגויה היה band shift.[68] אם השגיאה הייתה מתרחשת באללים פחות נדירים, סביר שהיא לא הייתה מתגלית.[69] גם כאשר נתונים פרופילים של אב, אם ובן משותף והפרופיל של הבן מלמד על אנומליה (למשל באתר מסוים חסר בו אלל של האם) קל לגלות את השגיאה.

פרופיל שגוי אינו מוביל בהכרח לשגיאה חיובית (false positive), שכן לצורך התרחשות שגיאה חיובית על הפרופיל השגוי להתאים גם לפרופיל של החשוד. לעומת זאת, יש לשגיאות אלה משמעות. אפילו טעות של 1%-2% בהפקת הפרופילים עלולה להוביל למסקנות מוטעות ב- linkage studies (מחקרים המנסים למצוא קשר בין מחלה מסוימת לבין מאפיינים גנטיים).[70] הספרות המקצועית דנה רבות בגורמי השגיאה, בהצעת שיטות להערכת שיעור הפרופילים השגויים ובהצעות להפחתת מספרם.[71] בין גורמי השגיאות יש להביא גם את שגיאות התוכנות המשמשות בהפקת הפרופיל.[72] גורמים נוספים הם זיהום או זיהום צולב, allelic dropout, false alleles, אי הקפדה על פרוטוקולי בדיקה, טעויות אנוש, אנומליות ביוכימיות ועוד.[73] ממחקרים נלמד כי השגיאה בהפקת הפרופיל הגנטי נעה בין 0.2%-15% לאתר נבדק, בהתאם למעבדה.[74] ברוב המעבדות, השגיאה היא 0.5%-1.0%. בתיקי עבודה בהם יש מעט DNA או DNA באיכות גרועה, שיעור השגיאה אף גבוה יותר.[75] בבדיקות מומחיות למעבדות פורנזיות באירופה נמצא ששיעור האתרים השגויים שהופקו בטכנולוגיות PCR בשנות האלפיים היה בין 0.4%-0.7% בהתאם לשנה.[76] עורכי המאמר לא כללו בסטטיסטיקת התוצאות שגיאות "פשוטות" כשגיאות אנוש. לו שגיאות אנוש היו מוכנסות לחישוב, אומדן השגיאה היה גבוה יותר. כאמור, אלה אחוזי השגיאה עבור אתר בודד, ומכאן כי אחוז השגיאה בפרופיל מלא המורכב ממספר אתרים גבוה יותר.[77] עולה מטבלאות 9 ו-10, כי התאמות ב-6 ואף 7 אתרים נפוצות מאד באוכלוסייה. לכן, יש להביא בחשבון גם כי פרופיל גנטי מלא שהופק מזירת הפשע ב-10 אתרים, שגוי במספר קטן של אתרים, בפרט כשחשוד מאותר

65 ראו הלפרט, לעיל ה"ש 7 הדן גם בקבוצות אנשים החולקים פרופיל גנטי משותף במספר משמעותי של אתרים.

66 E. SOBEL AT ALL. DETECTION AND INTEGRATION OF GENOTYPING ERRORS IN STATISTICAL GENETICS. 70 AM. J. HUM. GENET., 496, 496 (2002).

67 A. BONIN AL ALL, *HOW TO TRACK AND ASSESS GENOTYPING ERRORS IN POPULATION GENETICS STUDIES*. 13 MOLECULAR ECOLOGY 3261, 3261 (2004)

68 FORENSIC DNA EVIDENCE INTERPRETATION, לעיל ה"ש 26, בס' 5.3.5.

69 שם.

70 שם

71 POMPANON AT ALL, *GENOTYPING ERRORS: CAUSES, CONSEQUENCES AND SOLUTIONS*. 6(11), NATURE REV. GENET 847 (2005)

72 "genotype scoring software" ראו: Sobel at all לעיל ה"ש 66 בעמ' 496.

73 ראו טבלה, המסכמת גורמי שגיאה שונים אצל Pompanon at all, לעיל, ה"ש 71 בעמ' 849.

74 שם, בעמ 847.

75 שם.

76 S. RAND AT ALL, *THE GEDNAP BLIND TRIAL CONCEPT PART II. TRENDS AND DEVELOPMENTS*, 118, INT J LEGAL MED 83, 84 (2004).

77 אם ההסתברות לבצע שגיאה באתר מסוים היא e אזי ההסתברות שלא לבצע שגיאה היא 1-e. אם היינו מניחים אי תלות בשגיאה בין האתרים השונים (מה שלא בהכרח נכון) אזי ההסתברות שלא התרחשה השגיאה בהפקת פרופיל גנטי ב-n אתרים תהיה $(1-e)^n$. ומכאן ההסתברות שהתרחשה שגיאה בהפקת הפרופיל הגנטי תהיה $1-(1-e)^n$. אם נחשוב על חצי אחוז שגיאה לאתר בודד נקבל כי השגיאה עבור פרופיל גנטי המופק בעשרה אתרים תהיה $1-(1-0.005)^{10}=0.048$. כמעט 5% שגיאה. היינו, באחד מעשרים פרופילים גנטיים יש שגיאה בלפחות אתר אחד.

דרך מאגר, וזאת הראיה היחידה/עיקרית שנגדו.

גורם שגיאה נפוץ בהפקת הפרופיל הגנטי הוא זה המכונה allelic dropout. התופעה מתרחשת כאשר הבדיקה באתר הטרוזיגוט[78] מגלה, כתוצאה משגיאה בהפקת הפרופיל, רק אלל אחד מתוך השניים. התופעה מכונה לפעמים גם Null Allele או Silent Alleles. התופעה גורמת לאתר הטרוזיגוט להיראות כאתר הומוזיגוט[79]. יש המכנים את התופעה "הומוזיגוט כוזב" (false homozygote).

כאשר ה-DNA המשמש לבדיקה הוא בכמות מספיקה ובאיכות טובה, כפי שקורה כשהדגימה ניתנת מרצון ולא נלקחת מזירת פשע, התופעה אינה נפוצה.[80] כך למשל, מחקר שהשווה פרופילים שהופקדו מדגימות גנטיות באמצעות העֶרכָּה SGM PLUS, לפרופילים שהופקו מדגימות מאותם תורמים בערכה PowerPlex 16 גילה את התופעה ב-0.58% מהבדיקות.[81] כלומר ב-0.58% מהבדיקות בהן בערכה PowerPlex 16 התגלו שני אללים באתר מסוים, התגלה רק אלל אחד בערכה SGM PLUS. באחד המקרים, התופעה קרתה בשני אתרים באותו פרופיל. ההסבר המוצע הוא כי Primer-binding site mutations היא הגורם להשמטת האלל השני.[82] מאחר וה"פרימרים" שונים בין שתי הערכות, התופעה תקרה בערכה מסוג אחד ולא מהסוג האחר.

כאשר הבדיקה מבוצעת בכמות מעטה של חומר גנטי או חומר גנטי מאיכות לא טובה (כפי שקורה בדרך כלל בתיקי עבודה), התופעה נפוצה יותר.[83] למרות שיש הסברים שונים לתופעה בספרות, עדיין היא איננה מובנת לחלוטין.[84] לדוגמה, כאשר מדובר בתערובות אשר הפרופיל הגנטי של אחד התורמים (למשל הקורבן) ידוע אך בתערובת מופיע רק אלל אחד מבין שני האללים שלו. התופעה קיבלה הד בארה"ב לאחר הרשעתו של Robin Lovitt ברצח והוטל עליו עונש מוות.[85] אחת הראיות נגדו הייתה התאמה של אלל בודד מתוך שניים מדגימת דם שנלקחה מכלי הרצח.[86] הרשעתו התבססה למעשה על ההנחה השגויה במחלוקת כי באותו אתר קרתה התופעה של allelic dropout.[87] יום לפני הוצאתו להורג, הוא נחון ועונשו הומר למאסר עולם.[88]

רק בשנים האחרונות החלו להתפרסם בספרות הנחיות כיצד להתייחס לתערובות בהתחשב באפשרות של allelic dropout.[89] בין השאר, התייחסה ה-International Society of Forensic לפירוש של תערובות, וקבעה כי ככל שאלל אחד של חשוד, המופיע בתערובת הוא בעוצמה גבוהה יותר מעל סף הרעש, ואילו האלל השני אינו מופיע, מתחזקת האפשרות שהחשוד אינו התורם, עד מצב בו יש לקבוע כי ההסתברות שהוא התורם היא אפסית.[90] נראה כי המלצה זאת, אמנם עוזרת למניעת הרשעות שווא, אך במקרים מסוימים היא עדיין בעייתית ורחוקה מלפתור את בעיית ה-allelic dropout. בתיקי עבודה יש מקרים בהם אלל אחד של תורם ידוע לתערובת (כגון הקורבן) מופיע והוא מעל סף הרעש ואילו השני אינו מופיע. אין הגיון בשלילת האפשרות שאותו אלל שייך לתורם הידוע, מאחר ואנו מצפים לראות את האלל של התורם הידוע בתערובת. מקרים אלה לא תמיד זוכים להתייחסות מאחר וההתעניינות היא בעיקר בהימצאם

78 אתר הטרוזיגוט (heterozygote) הוא אתר המכיל שני אללים שונים.
79 אתר הומוזיגוט (homozygote) הוא אתר בו שני האללים זהים.
80 ראו סיכום תוצאות מחקרים רבים בנושא אצל
NIST, *Allele Dropout/Null Allele Reports, Results from Concordance Studies* available at:
http://www.cstl.nist.gov/div831/strbase/NullAlleles.htm

81 Vanderheyden al all, *Identification and sequence analysis of discordant phenotypes between AmpFlSTR SGM Plus and PowerPlex* 16. 121(4), *Int. J Legal Med.* 297 (2007)

82 Primer-binding site אלו האזורים השמורים באתר הנבדק אליהם מתחברים התחלים כדי לרבות את המקטע הנבדק. מוטציה באזור זה תמנע את התחברות התחלים, את האמפליפיקציה של המקטע ותגרום להשמטת האלל מהפרופיל הגנטי.

83 Pompanon at all לעיל, ה"ש 71, בעמ' 848.

84 Forensic DNA Evidence Interpretation, לעיל ה"ש 26, בס' 1.2.2.2.

85 William C. Thompson & Rachel Dioso-Villa, *Turning a Blind Eye to Misleading Scientific Testimony: Failure of Procedural Safeguards In a Capital Case* , 18 Alb. L.J. Sci. & Tech. 151 (2008)

86 שם, בעמ' 162, 165-168.

87 שם.

88 שם בעמ 188.

89 P.M. Schneider at all, *The German Stain Commission: Recommendations for the Interpretation of Mixed Stains*, 123 Int'l J. Legal Med. 1 (2009)

90 P. Gill al all, *'DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures.* 160 Forensic Sci Int. 90, 95-96 (2006)

או היעדרם של האללים של תורם לא ידוע, לצורך קביעת מבצע העבירה. אולם הם מלמדים, שיש אפשרות שאלל אחד יושמט מהפרופיל הגנטי באתר מסוים אף שעוצמת האלל השני שכן נצפה, היא מעל הרעש.

לאפשרות שאלל אחד יושמט מהפרופיל הגנטי גם כשהאלל השני הנצפה הוא מעל הרעש, יש כמובן משמעויות חשובות. שכן כאשר בפרופיל של מבצע העבירה, שהופק מהדגימה שהושארה בזירת הפשע, מופיעים אתרים הומוזיגוטים, אין לנו בטחון מלא כי הם אכן הומוזיגוטים. על כן אין לנו בטחון מלא כי הנאשם שהפרופיל הגנטי שלו מכיל את אותם אתרים הומוזיגוטים כמו אלה שהופקו ממבצע העבירה, הוא אכן מבצע העבירה. למשל, ייתכן מקרה בו בפרופיל של מבצע העבירה המתאים באופן מלא לחשוד ב-10 אתרים, יהיו 4 אתרים הומוזיגוטים. אולם כאשר אנו מביאים בחשבון מטבלאות 9 ו-10, כי צפויים עשרות מיליוני התאמות באוכלוסייה ב-6 אתרים מתוך ה-10 וכמעט מיליון התאמות ב-7 אתרים, וכאשר סביר מאד כי יהיו גם קבוצות אנשים גדולות התואמים באותם 6 אתרים ולא רק זוגות, יש הסבר חלופי להתאמה, לבד מההסבר כי החשוד הוא מבצע העבירה. יתכן כי מבצע העבירה אינו הומוזיגוט באותם האתרים אלא הוא הומזיגוט כוזב ב-4 האתרים הללו או בחלקם ועל כן הפרופיל הגנטי שלו אינו תואם לפרופיל של החשוד. הסבר זה מתחזק ממחקרו של Weir אשר הראה כי גם במעט אתרים הנוספים בהם לא נמצאה התאמה מלאה, צפויות התאמות באלל אחד מבין השניים.[91] הסבר חלופי זה להתאמה יכול להיות סביר במקרים בהם פרט לראיית ה-DNA אין ראיה משמעותית הקושרת את הנאשם לזירת הפשע ולפעמים יש גם ראיות לזכותו. לדבר יש משמעות גם במקרה בו קבוצת החשודים היא בני משפחה מורחבת בהם נהוגים נישואי קרובים. במשפחות כאלה התאמות ב-6 אתרים בין פרופילים של קרובי משפחה בדרגה של אחים, או אב וצאצאים אינן נדירות.[92]

לאור אפשרות תופעת allelic dropout, להלכה כל פרופיל גנטי שבו נצפים בכל אתר עד שני אללים, ואשר נחשב כיום כמגיע מתורם אחד, אמור להיות חשוד כתערובת. שכן אין ביטחון שלא מדובר בתערובת של שני תורמים, כאשר חלק מהאללים של התורמים אינם מופיעים. זאת, כנראה, בעיה אינהרנטית בטכנולוגית ה-DNA בימינו.

מקרה בו מופק מזירת הפשע פרופיל ובו רק מספר חלקי של אתרים, למשל 8 יוצר קושי. לפעמים, אתר אחד אינו תואם במדויק לזה של החשוד. ייתכן כי התביעה תטען מלכתחילה כי החשוד הוא אכן האשם ותנסה להסביר את היעדר האלל השני בכך שהוא נשמט, כפי שטען התובע במקרה Robin Lovitt. אך כאמור, מטבלאות 9 ו-10, עולה כי צפויים כמעט מיליון זוגות להם תואם גנטי ב-7 אתרים מתוך ה-10 ואשר צפויים להיות תואמים באלל אחד מבין השניים בשאר האתרים. מתוך זוגות אלה, רבים הפרופילים המשותפים למספר גבוה יותר של אנשים ולא רק לזוג. הסבר חליפי להשמטת האלל הוא כי לא החשוד ביצע את העבירה.

התעכבנו בפרוטרוט על בעיית ה- allelic dropout. אולם יש גורמים נוספים שעלולים להוביל לפרופיל שגוי, לאללים שגויים (mistaken allele) ואף לשגיאות חיוביות בבדיקה אשר מקורם המדויק לא התגלה אף בדיעבד.[93] לצערנו, אין סטנדרטים אובייקטיביים לקביעה מהו שיא (peak) אמיתי המייצג אלל ומהו אינו שיא אמיתי בתוצר הבדיקה.[94] יצרן הערכה SGM PLUS ממליץ במדריך למשתמש, כי שיא ייחשב כאלל אמיתי כאשר עוצמתו היא מעל סף הרעש המומלץ על ידו (150 RFU). אולם הוא גם קובע שמעבדות יכולות לעשות שימוש בניסיונן כדי לקבוע את הסף לעצמן.[95] במעבדות שונות ואפילו בבדיקות שונות באותה המעבדה, משתמשים בערכי סף שונים. מכאן, שקיימות מספר אפשרויות לניתוח

91 ראו Weir, לעיל ה"ש 13 והטקסט מעל.

92 מחברי המאמר נתקלו במקרה כזה בתיק שנדון בבית משפט בישראל בו המעבדה מצאה כי למבצע העבירה 4 אתרים הומוזיגוטיים. המדובר באותו מקרה באוכלוסיית חשודים מקרב משפחה אחת אשר בה נהוגים נישואי קרובים. השוואת 23 פרופילים גנטיים של אותם בני משפחה בינם לבין עצמם הניבה 4 התאמות בשישה אתרים ו-6 התאמות ב-5 אתרים. בחלק מההתאמות אף היו 16 אללים זהים מתוך 20.

93 כך במקרה הרצח של הפעוט ג'יידין לסקי אשר מתואר בדוח חוקר מקרי המוות באוסטרליה:
G. Johnstone, *Inquest into the death of Jaidyn Raymond Leskie*. 64-85 (2006) Coroners Case Number: 007/98.
סריקה במאגר נתונים בשנת 2003, הניבה התאמה בין שתי דגימות שנלקחו מבגדי הקורבן לבין פרופיל גנטי של אישה צעירה, מאותגרת מנטלית, אשר הדגימה שלה הייתה במאגר כתוצאה מכך שהיא הייתה קורבן אונס. הראיות החיצוניות ניקו את אותה צעירה מכל חשד למעורבות ברצח. אולם הוברר כי הדגימות הן ממקרה הרצח והן ממקרה האונס עובדו באותה מעבדה פחות או יותר באותה תקופת זמן (פברואר 1998) ועל כן, לאור הסתברות ההתאמה המקרית הנמוכה באותה מקרה, המסקנה ההגיונית היחידה של חוקר מקרי המוות אשר שמע מספר מומחים, הייתה שמקור ההתאמה הוא בזיהום צולב. כיצד שתי דגימות זוהמו ולא רק אחת, לא ידוע. אולם בדוח, חוקר מקרי המוות מצוין כי היו עוד מקרים שבהם התרחש זיהום אשר מקורו לא ידוע.

94 Thompson לעיל, ה"ש 85 בעמ' 169-168.

95 Kit User Manual, לעיל ה"ש 19, בעמ' 11-13

ואינטרפרטציה של תוצאות נתונות. תוצאות בדיקה אשר במעבדה אחת תפורש שאינה חד-משמעית (inconclusive) בהיבט הראיתי, ואולי אף כראיה מזכה, עלולה להתפרש במעבדה אחרת כראיה מרשיעה. אלה לדעתנו גורמים ההופכים את תוצאות הבדיקה במקרים מסוימים לסובייקטיביות, התלויות במעבדה בה הבדיקה מבוצעת ובמומחה המנתח אותה.

מסקנות והמלצות

מאמר זה עוסק בהתאמות החלקיות הרבות הצפויות להתקבל במאגרים ובאוכלוסייה כולה בישראל בין פרופילים גנטיים המופקים בערכה SGM PLUS בעשרה אתרים. תארנו מודל תיאורטי שיכול לשמש לחיזוי מספר ההתאמות החלקיות במאגר המכיל דגימות ממספר אוכלוסיות. את התחזיות שלנו ניתן לבדוק ישירות על ידי ספירת ההתאמות החלקיות במאגר הנמצא בידי הרשויות. הבדיקה תהווה מבחן הפרכה לתיאוריה הגנטית ובפרט תבחן את תקפות תיקון טטה. כבמדינות אחרות, אנו קוראים לגורמי האכיפה לספק לחוקרים מידע על המאגרים, תוך שמירה על פרטיות התורמים למאגר. בנוסף, הבאנו תחזית לגבי התאמות מלאות באוכלוסייה כולה. תחזית התאמות זאת מגיעה לשיעורים משמעותיים גם בשבעה אתרים. תארנו גם את התוצאות ממחקרו של Weir, הצופה התאמות חלקיות באלל אחד מתוך שניים, באותם האתרים בהם לא נמצאה התאמה מלאה. להתאמות אלה, גם שהן חלקיות משמעות רבה בתיקי עבודה, בהם בדרך כלל כמות ואיכות ה-DNA נמוכה. זאת לאור ההסתברות הגבוהה לקבלת פרופיל שגוי או חלקי בבדיקה בתנאים אלה, יחסית למצב בו הדגימות הן באיכות טובה ובכמות מספקת.

כאשר יש בסיס לחשש כי חסר מידע על אלל ואף על שני אללים באתר, ניתן לצמצם את הסיכון לזיהוי שגוי על ידי נקיטת הצעדים הבאים על ידי המעבדה:

א. בערכה SGM PLUS, המשמשת כיום להפקת פרופילים גנטיים על ידי רשויות התביעה בישראל, הגברת כל האתרים הנבדקים נעשית בו זמנית באותה מבחנה עצמה. שיטת עבודה זאת, לעומת הגברה של כל אתר בנפרד, היא זולה ויעילה ובדרך כלל היא מהימנה. אולם שיטה חסכונית זאת היא אחת הסיבות לקבלת פרופילים חלקיים, בעיקר בדגימות DNA באיכות ו/או בכמות נמוכה.[96] כאשר יש חשש לקבלת פרופיל חלקי באתר נתון ניתן לחזור על ההגברה באותו אתר בנפרד ולקבוע את הפרופיל הגנטי בו. שיטה זאת דורשת עבודה רבה יותר, אך היא מצמצמת את הסיכון לזיהוי שגוי.[97] כשיש חשש כי התקבל פרופיל גנטי חלקי בגלל מוטציה באתר ה"פריימינג" של אחד התחלים, ניתן לבדוק זאת על ידי שימוש בתחלים חלופיים.[98] צעדים אלה עשויים להעלות מעט את עלות הבדיקות, אולם הם מצמצמים את הסיכון לקבלת פרופיל שגוי.

ב. אחד האמצעים לקבוע, אם הימצאו של סוג אחד של אלל באתר נובע מחסרונו של האלל השני או מהימצאם של שני אללים זהים באותו אתר (הומוזיגוט), היא ניתוח עוצמת האות של התוצאה המתקבלת באתר בכלים כמותיים. באתר הומוזיגוטי בפרופיל גנטי צפויה להתקבל עוצמה כפולה של האלל היחיד בהשוואה לעוצמה הצפויה של אותו אלל בפרופיל הטרוזיגוטי. ניתן לבחון את האפשרות של אלל חסר על פי עוצמת האלל היחיד המתקבל.[99] פרוטוקולים של מעבדות מסוימות בעולם, דורשים, כי כדי לקבוע שפרופיל מסוים מגיע מתורם אחד, יש לראות כי העוצמות

96 ביצוע האמליפיקציה (ריאקציית ה-PCR) על כל האתרים בו זמנית באותה מבחנה מכונה - MULTIPLEX. כאשר נוקטים בשיטה זאת, בעיקר על DNA באיכות ו/או כמות נמוכה עלולות להתרחש אינטראקציות שונות בין התחלים (Primers) ובין התוצרים של האתרים השונים. כתוצאה מכך מתקבלת הגברה חלקית או חסרה בחלק המאתרים או באלל מסוים באתר. מאמצי המחקר הרבים שהוקדשו לחקר תופעה זו, עדיין לא הניבו הסבר מספק. ראו:
Bercovich D, Regev Z, Ratz T, Luder A, Plotsky Y, Gruenbaum Y. *Quantitative ratio of primer pairs and annealing temperature affecting PCR products in duplex amplification.* 27(4) Biotechniques: 762-4, 766-8, 770 (1999); Bercovich, D., Regev, Z., Ratz, T. and Plotsky, Y. 1997. The effect of quantitative ratio between primer pairs on PCR products in multi-band amplification. Modern Applications of DNA Amplification Techniques, Edited by Lassner et al., Plenum Press York: 83-89.

97 יתרה מכך, אם בערכה המסחרית היו נכללים, בנוסף למבחנה הכוללת תערובת של הפרימרים של כל האתרים, גם מבחנות הכוללות פרימרים לכל אתר בנפרד, ניתן היה לבצע את הפעולה בקלות יחסית. הדבר איננו קשה לביצוע שהרי הפרימרים לכל אתר מיוצרים בנפרד.

98 ניתן להכין לכל אתר, תחלים (פרימרים) חלופיים באמצעותם ניתן לחזור על הבדיקה (בשיטת MULTIPLEX או MONOPLEX בהתאם למקרה) במקרים הלא שכיחים בהם יש חשש לפרופיל חלקי כתוצאה ממוטציה באזור ה-PRIMING.

99 לא ניתן לערוך את ההשוואה באופן ישיר בין עוצמת האלל היחיד לבין עוצמת אללים באתרים אחרים באותו פרופיל גנטי, או בהשוואה לאלל באותו האתר בפרופיל גנטי אחר, אולם השוואה יחסית בין עוצמות האללים בתוך ובין אתרים ובתוך ובין פרופילים גנטיים, יכולה לסייע בהכרעה בשאלה אם מדובר באלל חסר או באתר הומוזיגוטי.

בתוך אתר ובין אתרים, מאוזנים.[100] למיטב ידיעתנו, במעבדות בישראל אין מבצעים ניתוח כמותי מובנה כדי לקבוע אם מדובר באתר הומוזיגוטי או באלל חסר.

ג. לאור עובדות אלה, יש המזהירים בפני תופעת אפקט הצופה (Observer Effects) בבדיקות DNA ובבדיקות פורנזיות בכלל.[101] משמעות המושג היא כי הצופה מפרש מידע שלפניו על פי ציפיותיו מהמידע. הסכנה היא שפרשנות התוצאה עלולה להיות מושפעת מן המידע הקודם שיש לעובד המעבדה.[102] מוצע, לפיכך, פרוטוקול לבדיקת DNA, המבוסס על "חשיפת מידע בשלבים" (Sequential Unmasking):[103] **א**. הפקת האללים מהפרופיל של מבצע העבירה. **ב**. קביעת מספר התורמים לפרופיל. **ג**. הערכת ההסתברות של השמטת אללים. **ד**. חישוב הסתברות ההתאמה המקרית באוכלוסיות שונות. שלבים אלה אמורים להיות מתועדים. מוצע להנהיג במעבדות סידורים שיבטיחו את הגינות התהליך ומניעת זליגת מידע פרט להכרחי. רק לאחר ביצוע השלבים שלעיל, יש לחשוף דגימות ייחוס לחשודים ולקבוע אם יש התאמה בין הדגימות רק על סמך המידע שהופק בשלבים הראשונים, כאשר המידע על החשוד, לא היה בנמצא.

במקרים רבים ראָיית ה-DNA עוזרת בפענוח פשעים. יש מקרים בהם חשוד אותר באמצעות ראיית ה-DNA ולאחר מכן נמצאו ראיות מחזקות.[104] אין אנו כופרים בפוטנציאל הזיהוי של ראיה זאת. אולם אסור להתעלם מן המוגבלות שבה ומהעובדה שהיא אינה חסינה בפני שגיאות. לעניין זה נבקש להפנות לדוח של פרופסור ריצ'ארד פיינמן, זוכה פרס נובל בפיסיקה, שעניינו חקירת אסון התרסקות מעבורת צ'לנג'ר. מדבריו עולה כי במהלך טיסות רבות התגלו בעיות באטמים בטילי ההנעה, בעיות אשר בסופו של דבר הובילו להתרסקות המעבורת. נבנָה אפילו מודל מתמטי אמפירי כדי להסביר את התופעה וכדי לקבוע מתי אפשר לאשר טיסה, למרות פגמים באטמים, זאת בהסתמך על "מקדמי בטחון". הצלחות קודמות בטיסות רק חיזקו את אמונת ראשי NASA כי המעבורת בטוחה. אולם הסיבות לבעיות באטמים לא הובנו לאשורן. כפי שפיינמן מציין בדוח שלו, הבעיות כלל לא היו אמורות להתעורר. כאשר למרות זאת תופעה שאינה אמורה להתרחש, בכל זאת מתרחשת, לא ניתן להסתמך על מקדמי ביטחון ניסיוניים, אלא אם גורמי התופעה מובנים במלואם:

> **"…The phenomenon of accepting for flight, seals that had shown erosion and blow-by in previous flights, is very clear. The Challenger flight is an excellent example. There are several references to flights that had gone before. The acceptance and success of these flights is taken as evidence of safety. But erosion and blow-by are not what the design expected. They are warnings that something is wrong. The equipment is not operating as expected, and therefore there is a danger that it can operate with even wider deviations in this unexpected and not thoroughly understood way. The fact that this danger did not lead to a catastrophe before is no guarantee that it will not the next time, unless it is completely understood. When playing Russian roulette the fact that the first shot got off safely is little comfort for the next. The origin and consequences of the erosion and blow-by were not understood. They did not occur equally on all flights and all joints; sometimes more, and sometimes less. Why not sometime, when whatever conditions determined it were right, still more leading to catastrophe?"**[105]

100 ראו הנחיה 3.1.1 בפרוטוקול Berkeley Dna Laboratory: Department of Justice, Bureau of Forensic Services, Berkeley Dna Laboratory. Protocol for the Extraction. and Quantitation of DNA From Case Evidence (6/17/1998)
ניתן לצפייה בכתובת:
http://www.nlada.org/Defender/forensics/for_lib/Documents/1120055953.78/CaliforniaDOJSOP.pdf

101 D. Krane, S. Ford, J. Gilder, K. Inman, A. Jamieson, R. Koppl, I. Kornfield, D. Risinger, N. Rudin, M. Taylor, W.C. Thompson. *Sequential unmasking: A means of minimizing observer effects in forensic DNA interpretation*. 53(4) Journal of Forensic Sciences, 1006-7 (2008).

102 שם.

103 שם.

104 מבין המקרים הבולטים נציין את פענוח רצח מעיין בן חורין בעזרת ראיית ה-DNA. ראו אתר ynet:
http://www.ynet.co.il/articles/0,7340,L-3367437,00.html כמתואר בגזר דינו, רוצח מעיין בן חורין לקח אחריות על מעשיו. פח 000402/07 מדינת ישראל נגד פלוני (אתר נבו 15/11/2007). ראו גם את פענוח הרצח של עו"ד ענת פלינר בעזרת ראיית DNA:
http://www.haaretz.co.il/hasite/spages/1068027.html

105 Richard P. Feynman, לעיל ה"ש 1.

כך גם בעניינו: תופעות כ-allelic dropout ואחרות אינן מוסברות דיין,[106] – עובדה שהיא אות אזהרה מפני זיהוי שגוי. גם הפער בין התחזיות למאגר באריזונה לנצפה בפועל מעורר פקפוק בתקפות הנחות העבודה ותיקון טטה. הצלחה בפענוח פשע בעבר, אינה מבטיחה מפני סכנת הרשעת חף מפשע בעתיד. כאשר ראיית ה-DNA מהווה ראיה יחידה או עיקרית, יש להביא בחשבון הסברים חלופיים להתאמה בבדיקת ה-DNA, לבד מזה כי הנאשם הוא אכן העבריין.

אין להמתין כי גם בישראל יתגלה בדיעבד כי ראיית DNA הובילה לזיהוי שגוי ולהרשעת שווא.[107] מחיר למידה מהניסיון, בַּתחומים בהם יש חשיבות לבטיחות עלול להיות גבוה מדי.[108] חובת הזהירות מחייבת נקיטה באמצעי מניעה שעלותם פחותה מתוחלת הנזק.[109] יש לוודא בחוק נקיטה באמצעי בטיחות בכל הקשור לשימוש בראיית ה-DNA ולראיות מדעיות אחרות, שאינם נהוגים כיום.[110] בין היתר, יש לקבוע כי פיתוח וייצור מכשור להפקת ראיות מדעיות, והשימוש בו, יהיו כפופים לכללי הנדסת בטיחות ולפיקוח מלא, כנהוג בתחום המכשור לאבחון וטיפול רפואי.[111]

נספח טכני

1. הקדמה

פרופילים גנטיים בישראל מופקים כיום ב-A=10 אתרים (חוץ מ כרומוזום המין). כל פרופיל הוא למעשה A זוגות מספרים. כל זוג מספרים מייצג את האללים באותו אתר. אנו מניחים כי איננו יכולים להבחין בין אלל שמקורו מהאב ובין אלל שמקורו מהאם. הקונבנציה היא שהאלל הקצר נרשם תחילה.

נניח כי באתר k של הפרופיל הגנטי ניתן למצוא n_k אללים שונים. נסמן את האללים השונים באתר k, על ידי זוג האינדקסים $\{i, j\}$ כאשר $i \in \{1,...n_k\}$ $j \in \{1,...n_k\}$.[112] טבלאות שכיחויות אללים באוכלוסייה u (למשל האוכלוסייה היהודית בישראל), מספקות את $p(i)$, ההסתברות למצוא את האלל i באתר k באוכלוסייה u. אנו מניחים כי הסתברות זאת מנורמלת כך ש:

106 ראה מקרה ג'יידין לסקי, לעיל ה"ש 93.

107 בארצות אחרות כבר התגלו הרשעות שווא המבוססות על סמך ראיית DNA. ראו הלפרט ופרדס, לעיל ה"ש 2 בעמ' 429-427.

108 Boaz Sangero & Mordechai Halpert, *From a Plane Crash to the Conviction of an Innocent Person: Why Forensic Science Evidence Should Be Inadmissible Unless It Has Been Developed As a Safety-Critical System*, 32 Hamline L. Rev. 2009) 65). גרסת המחברים למאמר זה ניתנת לצפייה: http://works.bepress.com/cgi/viewcontent.cgi?article=1000&context=dr_boaz_sangero; מרדכי הלפרט ובועז סנג'רו, הסכנה שבהרשעה על סמך נשיפה (על בדיקת הנשיפה המתיימרת לאתר נהיגה בשכרות), **הפרקליט – ספר דויד ווינר על המשפט הפלילי ואתיקה**, 313 (2009). ניתן לצפייה כאן:
http://www.hapraklit.co.il/_Uploads/dbsAttachedFiles/Sangero1.pdf

109 על נוסחתו של השופט לרנד האנד הקובעת כי התנהגותו של המזיק אינה סבירה כאשר שיעור הנזק, בהתחשב בהסתברות התרחשותו, עולה על ההוצאות למניעתו (B<Pl) ראו, למשל, ע"א 5604/94 **אוסמה חמד נ' מדינת ישראל**, דינים עליון, כרך סו 664, פס' 16 לפסק דינו של הש' ברק (2004). וראו גם קביעתו של כי "האדם הסביר אינו רק האדם היעיל. זהו גם האדם הצודק, ההוגן והמוסרי. זהו האדם הדואג לעצמו, לזולתו ולציבור, ואף כל אלה אינם משקפים את מלוא מורכבותו." שם. וראו גם פס' 3, 4, 6 של הש' ריבלין. שם.

110 ראו המקורות בה"ש 108.

111 שם. המחברים מתריעים כי יצרני מכשירים מהם מפיקים ראיות פורנזיות לא כפופים כלל לרגולציות ולכללי הנדסת בטיחות. הרשויות נותנות אימון עיוור ביצרן. היצרנים, אינם "מתנדבים" מיוזמתם לאמץ תקני בטיחות. דבר המוביל לשגיאות פשוטות (למשל בתוכנה) במכשירים פורנזיים הנמצאים בשימוש בפועל ואשר יכלו להימנע לו היצרנים היו מחויבים על ידי הרשויות לאמץ תקנים שכאלו. זאת בניגוד מוחלט לאשר נהוג בתחומים אחרים בחיינו אשר בטיחות בהם היא דבר "קריטי", כמו מכשירים רפואיים לאבחון רפואי, ציוד למטוסים או רישוי מכוניות. בתחומים אלו, הרשויות מפקחות על היצרנים מתוך מטרה למנוע את המחיר הכבד של למידה מהניסיון ומתוך מטרה שהמוצר יהיה בטיחותי כבר בשימוש ראשון. ערכה גנטית המשתמשת לריאקצית PCR לאבחון רפואי מחויבת ברגולציה קפדנית (אישר FDA בארה"ב או רשות מקבילה במדינה אחרת) בעוד שערכה בדיוק כזאת, לשימוש פורנזי פטורה מרגולציה. העובדה שאותה ערכה בדיוק אשר מספקת כדי להרשיע אדם ולגזור עליו בישראל מאסר עולם ובארה"ב אף להוצאתו להורג, אסורה לשימוש בתחום הרפואי מאחר ואין לה אישור FDA ועל כן לא הוכח שהיא בטוחה, היא בעינינו אנומליה חוקית ומשפטית. היא מלמדת כי המחוקק (לא רק בישראל) אינו רואה בהכרעה בעניין כל כך קריטי בחייו של אזרח, כמו הרשעתו או אי הרשעתו בעברה חמורה, חשיבות המצריכה שימוש באמצעי בטיחות כמו אלו הננקטים למניעת אבחון רפואי שגוי. ראו דברי יצרן ערכת SGM PLUS (חברת Applied Biosystems), אשר בחר בשורות התוכן הראשונות במדריך למשתמש בערכה, מיד אחרי עמוד השער ואחרי משפט פתיחה בדבר זכויות יוצרים, להתריע בכתב מודגש בפני השימוש בערכה לצרכים רפואיים:
"For Research, Forensic and Paternity Use Only. Not for use in diagnostic procedures."
ראו Kit User Manual לעיל ה"ש 19, בעמ' 2 (ההדגשה במקור).

112 בסימון זה האללים אינם מיצגים את מספר החזרות של קבוצת בסיסים מסוימת באתר מסוים ב-DNA, אלא את המספר הסידורי בלבד של חזרה מסוימת, כאשר האלל הקצר ביותר נחשב 1 והאלל הארוך ביותר יסומן כ- n_k. למשל אם האלל הקצר ביותר באתר מסוים הוא 11, אזי בסימונים שלנו הוא יקרא אלל 1. אם מספר החזרות הבא שנמצא הוא 13, אזי זה יסומן כאלל-2 וכולי.

$$\sum_{i=1}^{n_k} p(i) = 1 \quad (1)$$

חוק HW מניח אי תלות בין שני האללים באותו אתר ולכן ניתן להשתמש בכלל המכפלה ההסתברותי כדי לחשב את ההסתברות למצוא את זוג האללים $\{i, j\}$ באתר k:

$$P(k)_{\{i,j\}} = \begin{cases} 2 \times p(i) \times p(\mathrm{j}), \mathrm{i} \neq \mathrm{j} \\ p(i)^2, \mathrm{i} = \mathrm{j} \end{cases} \quad (2)$$

כאשר $1 \leq i \leq j \leq n_k$.

לשם הנוחיות אנו מניחים כי n_k קבוע לכל האוכלוסיות השונות. כלומר אנו מניחים כי n_k כולל בתוכו את מספר האללים האפשריים באתר k בכל האוכלוסיות. אם באוכלוסייה מסוימת (נניח האוכלוסייה הערבית) אלל מסוים m, אינו מופיע, אנו מציבים $p(m) = 0$.

ממשוואה 1 נובע כי:

$$\sum_{i=1, j \geq i}^{n_k} P(k)_{\{i,j\}} = 1 \quad (3)$$

בהנחת קיומו של Linkage Equilibrium, היינו שאין תלות בין האללים באתרים השונים, ניתן לחשב את הסתברות ההתאמה המקרית עבור פרופיל גנטי המיוצג על ידי קבוצת זוגות המספרים $\{i_k, j_k\}$ $1 \leq k \leq A$ באופן הבא:

$$rmp_{\{i_k, j_k\}} = \prod_{k=1}^{A} P(k)_{\{i_k, j_k\}} \quad (4)$$

2. חישוב מספר ההתאמות הצפוי במאגר אחיד הכולל N פרופילים גנטיים המתפלגים על פי שכיחות אללים באוכלוסייה אחת.[113]

נתון לנו מאגר המכיל N פרופילים גנטיים שנדגמו מאוכלוסייה מסוימת (לדוגמה, מהאוכלוסייה היהודית). נרצה לחשב את מספר ההתאמות המלאות (זהות מלאה בכל האתרים של שני פרופילים גנטיים שונים) הצפוי במאגר. נסמן את הסתברות ההתאמה המקרית של פרופיל i ב- rmp^i כאשר $1 \leq i \leq N$. האינדקס i מייצג פרופיל גנטי אשר האללים שלו מיוצגים על ידי הקבוצה $\{i_k, j_k\}$. מספר ההתאמות המלאות הצפוי במאגר יהיה:

$$< n >= \frac{1}{2} \sum_{i=1}^{N} (N-1) \times rmp^i = \frac{N \times (N-1)}{2} \times \overline{rmp} \quad (5)$$

כאשר

$$\overline{rmp} = \frac{\sum_{i=1}^{N} rmp^i}{N} \quad (6)$$

הביטוי $\overline{rmp}$ המופיע במשוואה (6), מייצג למעשה את הסתברות ההתאמה המקרית הממוצעת של הפרופילים במאגר, או כפי שהוא מכונה לעיתים Probability of Identity. זאת ההסתברות **ששני אנשים אקראיים** באוכלוסייה יחלקו את אותו פרופיל גנטי. הסתברות זאת שונה מהסתברות ההתאמה המקרית הרגילה שכן האחרונה מתייחסת להסתברות כי אדם אקראי באוכלוסייה יתאים לפרופיל מסוים הנתון בידינו. בהנחה כי התפלגות האללים בין הפרופילים הגנטיים במאגר זהה להתפלגות האללים כפי שנדגמו באוכלוסייה, ואשר נתונות בטבלאות שכיחות האללים, ניתן לחשב את הערך התיאורטי של $\overline{rmp}$ שבמשוואה (6) מאותן טבלאות. תחילה, יש לחשב את הסתברות ההתאמה המקרית התיאורטית הממוצעת באתר k מסוים. באתר k יתכנו האללים i ו- j כאשר $1 \leq i \leq j \leq n_k$. מכאן, שיש l_k קומבינציות שונות לקבלת זוג אללים i ו-j באותו אתר כאשר:

$$l_k = n_k + 0.5 \times n_k \times (n_k - 1) \quad (7)$$

113 פרק זה לנספח כבר תואר על ידי הלפרט, לעיל ה"ש 7.

ההסתברות של כל קומבינציה אפשרית של אללים להופיע בפרופיל גנטי נתונה על ידי הביטוי $P(k)_{\{i,j\}}$ במשוואה (2). לכן הסתברות התאמה מקרית ממוצעת באתר k נתונה כערך התצפית של הביטוי $P(k)_{\{i,j\}}$ באופן הבא[114]:

$$\overline{rmp(k)} \equiv \overline{P(k)} = \sum_{i,j,i\le j}^{n_k} [P(k)_{\{i,j\}}]^2 = 2\times\left(\sum_{i=1}^{n_k} p(i)^2\right)^2 - \sum_{i=1}^{n_k} p(i)^4 \qquad (8)$$

כעת נניח שיש לנו שני אתרים k_1 ו- k_2 עם l_{k1} ו- l_{k2} בהתאמה, קומבינציות אפשריות של אללים, כנתון במשוואה (7). המספר הכללי של קומבינציות של אללים בשני האתרים גם יחד יהיה $l_{k1}\times l_{k2}$. אבל

$$\sum_{i_1,j_1,i_1\le j_1}^{n_{k1}}\sum_{i_2,j_2,i_2\le j_2}^{n_{k2}}[P(k_1)_{\{i_1,j_1\}}]^2\times[P(k_2)_{\{i_2,j_2\}}]^2 = \left(\sum_{i_1,j_1,i_1\le j_1}^{n_{k1}}[P(k_1)_{\{i_1,j_1\}}]^2\right)\times\left(\sum_{i_2,j_2,i_2\le j_2}^{n_{k2}}[P(k_2)_{\{i_2,j_2\}}]^2\right) \qquad (9)$$

ומכאן שניתן להשתמש בכלל המכפלה עבור הסתברות התאמה מקרית ממוצעת בשני אתרים. שימוש באינדוקציה יניב כי הדבר נכון גם לכל מספר של אתרים. מכאן נקבל כי:

$$\overline{rmp} = \prod_{k=1}^{A}\overline{P(k)} = \prod_{k=1}^{A}\overline{rmp(k)} \qquad (10)$$

3. מספר ההתאמות המלאות הצפוי במאגר הכולל N פרופילים גנטיים כאשר מקור התורמים למאגר הוא מכמה אוכלוסיות

נניח כי המאגר כולל N פרופילים גנטיים שנדגמו מ-M אוכלוסיות בעלות התפלגות אללים שונה (למשל, מהאוכלוסייה היהודית והערבית - M=2). משמעות המעשית היא כי התפלגות האללים בכל אוכלוסייה נתונה על ידי טבלת הסתברויות נפרדת. מספר הפרופילים בכל אוכלוסייה נתון על ידי $S(v)$ כאשר:

$$\sum_{v=1}^{M} S(v) = N \qquad (11)$$

בחישוב מספר ההתאמות הצפוי נתייחס לשתי אפשרויות. האחת, התאמה פנימית (Intra-Match להלן IM) המתייחסת לאפשרות התאמות **בתוך** קבוצת אוכלוסייה (למשל, פרופיל מתורם באוכלוסייה היהודית זהה לפרופיל של תורם אחר באוכלוסייה). האפשרות השנייה היא התאמה חיצונית (External Match להלן EM) כאשר פרופיל של תורם מאוכלוסייה אחת תואם לפרופיל של תורם מאוכלוסייה אחרת.

מספר התאמות הכולל הצפוי במאגר $< n^M >$ יהיה הסכום של מספר ההתאמות הפנימיות הצפוי $< IM^M >$ עם מספר ההתאמות החיצוניות הצפוי $< EM^M >$:

$$< n^M >=< IM^M > + < EM^M > \qquad (12)$$

כאשר האינדקס העילי M בה לציין שהמאגר מכיל M קבוצות אוכלוסייה.

---

114 ערך התצפית של n משתנים yi, המתפלגים על פי שכיחות $p(y_i)$ הוא $< y >= \frac{\sum_{i=1}^{n} y_i p(y_i)}{\sum_{i=1}^{n} p(y_i)}$.

אם מניחים כי השכיחות מנורמלת $\sum_{i=1}^{n} p(y_i) = 1$ ניתן להשמיט את גורם הנרמול מערך התצפית כך ש- $< y >= \sum_{i=1}^{n} y_i p(y_i)$.

משוואה (8) מתקבלת כאשר מבחינים כי המשתנים הינם בעצם השכיחויות עצמן $p(y_i)$, היינו $< p >= \sum_{i=1}^{n} p(y_i)\times p(y_i) = \sum_{i=1}^{n} p(y_i)^2$

ראו: Bruce S. Weir, Matching and Partially-Matching DNA Profiles. 49(5) J. Forensic Sci. 1009, 1014 (2004)

על ידי שימוש במשוואה (5) לכל קבוצת אוכלוסייה בנפרד וסיכום על כל האוכלוסיות, נוכל לקבל את מספר ההתאמות הפנימיות הצפוי במאגר:

$$< IM^M >= \sum_{v=1}^{M} \frac{S(v)\times(S(v)-1)}{2}\times\overline{rmp_v} \quad (13)$$

כאשר האינדקס v המופיע בביטוי $\overline{rmp_v}$ במשוואה לעיל, מציין כי בחישוב הסתברות הזהות $\overline{rmp}$ הנתונה במשוואה (10), על ידי שימוש במשוואות (6) ו-(8), יש להשתמש בטבלאות התפלגויות אללים של אוכלוסייה v.

כדי לחשב את מספר ההתאמות החיצוניות בין שתי אוכלוסיות שונות, מניחים כי פרופיל גנטי i שייך לתורם מאוכלוסייה u. rmp_v^i היא הסתברות של הפרופיל i להתאים באופן מקרי לפרופיל אשר התורם שלו שייך לאוכלוסייה אחרת - v. זאת אותה הסתברות התאמה מקרית הנתונה במשוואה (4), אך היא מחושבת מטבלאות אללים של אוכלוסייה v. מספר ההתאמות הצפוי של פרופיל i לפרופילים גנטיים מאוכלוסייה v יהיה $rmp_v^i \times S(v)$. ומכאן, סיכום על כל הפרופילים i באוכלוסייה u תיתן את מספר ההתאמות החיצוניות הצפוי בין פרופילים מאוכלוסייה u לפרופילים מאוכלוסייה v:

$$< n(u \to v) >= \sum_{i=1}^{S(u)} rmp_v^i \times S(v) = S(u)\times S(v)\times\overline{rmp_v^u} \quad (14)$$

כאשר $\overline{rmp_v^u}$ הוא המקרית הסתברות הזהות של פרופילים מאוכלוסייה u להתאים לפרופילים מאוכלוסייה v:

$$\overline{rmp_v^u} = \prod_{k=1}^{A}\overline{rmp_v^u}(k) \quad (15)$$

וכאשר:

$$\overline{rmp_v^u}(k) = \sum_{i,j,i\le j}^{n_k}[p^u(k)_{\{i,j\}}\times p^v(k)_{\{i,j\}}] \quad (16)$$

כאשר $p^f(k)_{\{i,j\}}$ במשוואה לעיל, מחושב על פי משוואה (2) וכאשר התוספת של האינדקס f באה לציין כי בחישוב גודל זה, יש לעשות שימוש בטבלאות אללים המתאימות לאוכלוסייה f.

ממשוואה 16 עולה כי $\overline{rmp_v^u} = \overline{rmp_u^v}$ (סימטריה בין אוכלוסייה u לאוכלוסייה v). לכן, תוך התייחסות למשוואות 14 ו-15 ניתן לראות כי מספר ההתאמות החיצוניות הצפוי בין אוכלוסייה v לאוכלוסייה u זהה למספר ההתאמות החיצוניות הצפוי בין אוכלוסייה u לאוכלוסייה v (כפי שאמור להיות).

מספר ההתאמות החיצוניות הצפוי בין M האוכלוסיות נתון אם כך בביטוי הבא:

$$< EM^M >= \sum_{u,v,u<v}^{M}(S(u)\times S(v)\times\overline{rmp_v^u}) \quad (17)$$

איחוד משוואות 12, 13 ו-17 נותן את מספר התאמות הכולל הצפוי במאגר $< n^M >$:

$$< n^M >= \sum_{v=1}^{M}\left(\frac{S(v)\times(S(v)-1)}{2}\times\overline{rmp_v}\right)+\sum_{u,v,u<v}^{M}\left(S(u)\times S(v)\times\overline{rmp_v^u}\right) \quad (18)$$

4. חישוב מספר ההתאמות החלקיות הצפוי ב w אתרים כלשהם מתוך n במאגר אחיד הכולל N פרופילים גנטיים.

בחלק זה אנו מניחים כי המאגר כולל N פרופילים גנטיים המופקים מ-n=10 אתרים. הנחה נוספת היא כי מקור הפרופילים הוא בתורמים המשתייכים לאוכלוסייה אחת בלבד (למשל היהודית). התאמה חלקית של פרופילים גנטיים ב-w אתרים מתוך n משמעותה כי שני פרופילים תואמים ב-w אתרים ולא תואמים ב-(n-k) האתרים האחרים.

כעת, ישנן $L \equiv C_w^n = \frac{n!}{(n-w)!\times w!}$ אפשרויות לבחור w אתרים מתוך n.

נסמן במספר הסידורי l אפשרות אחת שכזאת כאשר $1 \le l \le L$. נגדיר את הווקטור $w^l \equiv \{k_1^l, k_2^l, ...k_w^l\}$ כמייצג את w האתרים המסוימים מתוך n, עבור האפשרות l, כאשר $1 \le k_s^l \le n$. הסתברות ההתאמה המקרית החלקית של פרופיל גנטי מסוים שיתאים לפרופיל אחר באותם w^l אתרים ולא יתאים בשאר (n-k) האתרים הנוספים נתונה בביטוי הבא:

$$Q(n, w^l) = \prod_{s=1}^{w} (p(k_s^l) \times \prod_{s=w+1}^{n} \left(1 - p(k_s^l)\right) \quad (19)$$

ממשוואה (5) ומהטקסט המסביר אותה, ניתן לראות כי ערך התצפית של מספר ההתאמות החלקיות בין פרופילים במאגר באותם w^l **אתרים מסוימים**, הוא:

$$< n_{w_l} >= \frac{N \times (N-1)}{2} \times \overline{Q(n, w^l)} \quad (20)$$

כאשר:

$$\overline{Q(n, w^l)} = \prod_{s=1}^{w} \overline{p(k_s^l)} \times \prod_{s=w+1}^{n} \left(1 - \overline{p(k_s^l)}\right) \quad (21)$$

וכאשר $\overline{p(k_s^l)}$ ניתן במשוואה (8)

מכאן, שהמספר הצפוי של התאמות חלקיות בכל w אתרים מתוך n יהיה:

$$< n_w >= \frac{N \times (N-1)}{2} \times \sum_{l=1}^{L} \overline{Q(n, w^l)} \quad (22)$$

5. חישוב מספר ההתאמות החלקיות הצפוי ב w אתרים כלשהם מתוך n במאגר לא אחיד הכולל N פרופילים גנטיים כאשר מקור התורמים למאגר הוא ממספר אוכלוסיות שונות.

בחלק זה אנו מניחים כי המאגר כולל N פרופילים גנטיים המופקים מ-n=10 אתרים. בניגוד לסעיף הקודם, ההנחה היא כי מקור הפרופילים במאגר הוא בתורמים המשתייכים ל-M אוכלוסיות שונות. בדומה לדיון בסעיף 3 לנספח זה נקבל כי:

$$< n_w^M >=< IM_w^M > + < EM_w^M > \quad (23)$$

כאשר $< n_w^M >$ במשוואה שלעיל הוא מספר ההתאמות הצפוי ב-w אתרים כלשהם מתוך n מתוך N הפרופילים הגנטיים. הביטוי $< IM_w^M >$ המופיע במשוואה לעיל, מתייחס לסיכום מספר ההתאמות **הפנימיות** החלקיות ב-w אתרים כלשהם מתוך n, הצפוי בקרב כל אחת מהאוכלוסיות התורמות למאגר (למשל בין יהודים ליהודים ובין ערבים לערבים). הביטוי $< EM_w^M >$ המופיע במשוואה לעיל מתייחס לסיכום מספר ההתאמות **החיצוניות** (למשל בין יהודים לערבים) החלקיות ב-w אתרים כלשהם מתוך n, הצפוי בקרב כל אחת מאוכלוסיות התורמות למאגר.

שילוב משוואות 13 ו-22 יאפשר לנו לחשב את $< IM_w^M >$:

$$< IM_w^M >= \sum_{v=1}^{M} \left[\frac{S(v) \times (S(v)-1)}{2} \times \sum_{l=1}^{L} \overline{Q(n, w_v^l)} \right] \quad (24)$$

כאשר $\overline{Q(n, w_v^l)}$ מחושב על פי משוואה (21), וכאשר תוספת האינדקס v מציינת כי בחישוב הביטוי $\overline{p(k_s^l)}$ המופיע במשוואה (8) יש להשתמש טבלת שכיחויות אללים הרלוונטית לאוכלוסייה v.

כדי לחשב את הגודל $< EM_w^M >$ נחליף את משוואה (17) בביטוי הבא:

$$< EM_w^M >= \sum_{u,v,u<v}^{M} [S(u) \times S(v) \times \sum_{l=1}^{L} \overline{Q_v^u(n,w^l)}] \quad (25)$$

כאשר:

$$\overline{Q_v^u(n,w^l)} = \prod_{s=1}^{w} \overline{P_v^u(k_s^l)} \times \prod_{s=w+1}^{n} (1 - \overline{P_v^u(k_s^l)}) \quad (26)$$

וכאשר:

$$\overline{P_v^u(k)} = \sum_{i,j,i\le j}^{n_k} [p^u(k)_{\{i,j\}} \times p^v(k)_{\{i,j\}}] \quad (27)$$

שילוב משוואות 23, 24, 25, נותן כי

$$< n_w^M >= \sum_{v=1}^{M} [\frac{S(v) \times (S(v)-1)}{2} \times \sum_{l=1}^{L} \overline{Q(n,w_v^l)}] \quad (28)$$
$$+ \sum_{u,v,u<v}^{M} [S(u) \times S(v) \times \sum_{l=1}^{L} \overline{Q_v^u(n,w^l)}]$$

הצבת המספרים מבוצעת בעזרת תוכנת מחשב.

6. תיקון לתת-אוכלוסיות

את התיקון לתת-אוכלוסיות ביצענו על סמך מודלים מתאימים המתוארים בספרות.[115] ההנחה היא כי האללים, ששכיחותם נתונה רק באוכלוסייה הכללית, אינם בלתי תלויים בתת-אוכלוסייה. על כן, משוואה 2 אינה מתארת באופן נכון את שכיחות צמד אללים מסוים באתר מסוים בתת- האוכלוסייה. עבור תת-אוכלוסייה משוואה 2 מוחלפת במשוואה הבאה:[116]

$$P(k)_{\{i,j\}} = \begin{cases} 2 \times p(i) \times p(\mathrm{j}) \times (1-\theta) & \mathrm{i} \neq \mathrm{j} \\ p(i)^2 + \theta \times p(i) \times (1-p(i)) & \mathrm{i} = \mathrm{j} \end{cases} \quad (29)$$

כאשר θ מייצג את הסטייה משיווי משקל HW. התיקון מעלה את ההסתברות של האתר להיות הומוזיגוט אולם מפחית את ההסתברות של האתר להיות הטרוזיגוט. קל להיווכח כי גם לאחר התיקון, הערך $P(k)_{\{i,j\}}$ הנתון במשוואה (29) מנורמל ומשוואה (3) תקפה.

מצד שני, בחישוב הסתברויות ההתאמה המקרית בתיקי עבודה ובחישוב הסתברות הזהות, אין משתמשים באופן ישיר בערכים הנתונים במשוואה (29). המצא אלל מסוים בפרופיל הגנטי, מגדיל את ההסתברות למצוא אלל נוסף כזה. זאת, מאחר ובתת-אוכלוסייה יש תלות בין האללים השונים, ועל כן, כאשר נצפה אלל אחד מסוג מסוים, עולה גם ההסתברות כי למישהו אחר באוכלוסייה יש אלל זהה. לכן, נטען בספרות כי אין זה נכון להשתמש בנוסחה (29) בחישוב ההסתברות ששני אנשים שונים בתת-האוכלוסייה חולקים את אותו פרופיל גנטי. בספרות העוסקת בתיקון טטה מקובל להתייחס למשוואה (29) כמשוואה המתארת את ההסתברות לקבלת צמד אללים מסוים באתר מסוים **בפרופיל אחד** באוכלוסייה.[117]

115 ראו Weir, לעיל ה"ש 13; Weir, לעיל ה"ש 114;
David Balding, Weight-of-Evidence for Forensic DNA Profiles. John Wiley & Sons.

116 ראו משוואות 4.4a ו-4.4b בע' 102 בדוח NRC-II National Research Council, The Evaluation of Forensic DNA Evidence (1996) 102

117 Forensic DNA Evidence Interpretation, לעיל ה"ש 26, בס' 3.3.3.

לצורך חישוב הסתברות ההתאמה בתת-אוכלוסייה מניחים כי ההסתברות למצוא אלל A_i באתר מסוים בה, בהינתן שהוא נצפה m פעמים בעבר מתוך n תצפיות, נתון בביטוי הבא:[118]

$$P_{m,n}A_i = \frac{m\times\theta + p_i\times(1-\theta)}{1+(n-1)\times\theta} \qquad (30)$$

כאשר p_i מייצג את שכיחות האלל A_i כפי שנדגמה באוכלוסיית הייחוס. עבור n = 0 ו- m = 0, נקבל כי ההסתברות למצוא אלל זה כאשר לפני כן לא ניסינו ולא מצאנו אלל זה היא p_i, בדיוק כבאוכלוסייה הכללית. אם נפעיל את המשוואה פעם אחת עבור n = 0 ו- m = 0 ואז נחזור ונפעיל אותה עבור m=1, n=1 ונכפיל ערכים אלה, נקבל את ההסתברות למצוא אתר הומוזיגוטי באוכלוסייה, כפי שנתון במשוואה (29). אם נבצע את החישוב פעם אחת עבור n = 0 ו- m = 0, ופעם נוספת עבור m=0, n=1, ונכפיל ערכים אלה נקבל את ההסתברות למצוא אתר הטרוזיגוטי באוכלוסייה. הוספת פקטור 2 לחישוב, כתוצאה מכך שאיננו יודעים איזה אלל מקורו באב ואיזה באם, מוביל לתוצאה הנתונה במשוואה (29).

כעת, בתיק עבודה, נניח כי ידוע שמבצע העבירה הוא הומוזיגוט באתר מסוים (בעל אללים $\{A_i, A_i\}$ באותו אתר) וכך גם החשוד. אנו מעוניינים לחשב את ההסתברות שלאדם אקראי אחר יהיו אללים זהים באותו אתר. מאחר ויש בידינו פרופיל של מבצע העבירה נפעיל את משוואה (30) עבור ערכי m=2, n=2, ולאחר מכן עבור ערכי m=3, n=3 ונכפיל את שתי ההסתברויות. נקבל את נוסחת העבודה הרגילה בתיקי עבודה לחישוב הסתברות התאמה מקרית עבור אתרים הומוזיגוטים:[119]

$$P(A_i, A_i \mid A_i, A_i) = \frac{\left(2\times\theta + (1-\theta)\times p(A_i)\right)\times\left(3\times\theta + (1-\theta)\times p(A_i)\right)}{\left(1+\theta\right)\times\left(1+2\times\theta\right)} \qquad (31)$$

כאשר $P(A_i, A_i \mid A_i, A_i)$, היא ההסתברות שלאדם אקראי בתת-האוכלוסייה אליה החשוד משתייך, יש האללים $\{A_i, A_i\}$ באתר מסוים, בהינתן שגם לחשוד יש אותם האללים.

באותו אופן, עבור אתר הטרוזיגוטי נפעיל את הנוסחה קודם עבור m = 1, n = 2 ואחר כך עבור m = 1, n = 3 ואחרי הוספת פקטור 2, נקבל את נוסחת העבודה הרגילה בתיקי העבודה לחישוב הסתברות התאמה מקרית עבור אתרים הטרוזיגוטים:[120]

$$P(A_i, A_j \mid A_i, A_j) = 2\times\frac{\left(\theta + (1-\theta)\times p(A_i)\right)\times\left(\theta + (1-\theta)\times p(A_j)\right)}{\left(1+\theta\right)\times\left(1+2\times\theta\right)} \qquad (32)$$

כאשר $P(A_i, A_j \mid A_i, A_j)$, היא ההסתברות שלאדם אקראי בתת-האוכלוסייה אליה החשוד משתייך, יש האללים $\{A_i, A_j\}$ באתר מסוים, בהינתן שגם לחשוד יש אותם האללים.

אולם משוואות (31) ו-(32) אינן ישימות עבור מקרה בו אנו מחפשים התאמות במאגר מאחר ואין בידינו מראש פרופיל אחד אותו אנו משווים לאחר. לכן יש להביא בחשבון גם את ההסתברות למצוא את הפרופיל הראשון.[121] הסתברות זאת שונה בין התאמות פנימיות להתאמות חיצוניות.

עבור התאמות פנימיות, ועבור אתרים הומוזיגוטים וכדי לחשב את ההסתברות ששני פרופילים באוכלוסייה יהיו תואמים באתר אחד, יש להפעיל את נוסחה (30) 4 פעמים פעם על m=0, n=0, אז על m=1, n=1, אז על m=2, n=2, ואז על m=3, n=3. אז יש להכפיל ערכים אלה אחד בשני כדי לקבל[122]:

118 ראו Weir, לעיל ה"ש 13.
119 ראו משוואה 4.10a, בדוח NRC-II לעיל ה"ש 116, בעמ' 114.
120 ראו משוואה 4.10b, בדוח NRC-II לעיל ה"ש 116, בעמ' 115.
121 ראו Weir, לעיל ה"ש 114.
122 ראו ביטוי זה בנספח אצל Weir, לעיל ה"ש 114.

$$P(A_i, A_i, A_i, A_i) = p(A_i) \times \left(\theta + (1-\theta) \times p(A_i)\right) \times \frac{\left(2 \times \theta + (1-\theta) \times p(A_i)\right) \times \left(3 \times \theta + (1-\theta) \times p(A_i)\right)}{(1+\theta) \times (1+2 \times \theta)} \quad (33)$$

כאשר $P(A_i, A_i, A_i, A_i)$ מייצג את ההסתברות למצוא שני פרופילים גנטיים בתת-אוכלוסייה, להם אתר הומוזיגוטי זהה המכיל את האללים המסוימים $\{A_i, A_i\}$.

באותו אופן ועבור אתר הטרוזיגוט נקבל כי:[123]

$$P(A_i, A_j, A_i, A_j) = \frac{4p(A_i)p(A_j)(1-\theta)\left(\theta + (1-\theta)p(A_i)\right)\left(\theta + (1-\theta)p(A_j)\right)}{(1+\theta)(1+2\times\theta)} \quad (34)$$

כאשר $P(A_i, A_j, A_i, A_j)$ מייצג את ההסתברות למצוא שני פרופילים גנטיים בתת-אוכלוסייה, להם אתר הטרוזיגוטי זהה המכיל את האללים המסוימים $\{A_i, A_j\}$.

מכאן, משוואה (8) תוחלף בביטוי הבא:[124]

$$\overline{rmp(k)} \equiv \overline{P(k)} = \sum_{i,j,i \le j}^{n_k} P(A_i, A_j, A_i, A_j) \quad (35)$$

כאשר בסכימה במשוואה לעיל, יש להשתמש עבור מקרים בהם $i=j$ בביטוי המופיע במשוואה (33) ועבור מקרים בהם $i \neq j$ יש להשתמש בביטוי המופיע במשוואה (34). שינוי זה יחול כמובן גם על משוואה (21) העושה שימוש במשוואה (8) עבור התאמות חיצוניות החישוב הוא שונה. זאת מאחר והמצאות אללים מסוימים בתת-אוכלוסייה מסוימת של אוכלוסיית הייחוס (למשל האוכלוסייה היהודית) אינה מלמדת דבר על הימצא אותם אללים בתת-אוכלוסייה באוכלוסיית יחוס אחרת (למשל האוכלוסייה הערבית). לכן, נפעיל את משוואה (30) באוכלוסיית יחוס אחת u, פעם אחת עבור $m = 0$ ו- $n = 0$ ואז נפעיל אותה שוב עבור $m=1, n=1$. נכפיל ערכים אלה אחד בשני כדי לקבל את ההסתברות לקבל צמד אללים באוכלוסייה u. נחזור על התהליך עבור אוכלוסיית יחוס v. הכפלת הערכים משתי אוכלוסיות הייחוס תיתן כי:

$$P(A_i^u, A_j^u, A_i^v, A_j^v) = \left\{ \begin{array}{l} 4 \times p^u(i)^2 \times p^v(\mathrm{j})^2 \times (1-\theta)^2, \mathrm{i} \neq \mathrm{j} \\ \left(p^u(i)^2 + \theta \times p^u(i) \times (1-p^u(i))\right) \\ \times \left(p^v(i)^2 + \theta \times p^v(i) \times (1-p^v(i))\right), \mathrm{i} = \mathrm{j} \end{array} \right\} \quad (36)$$

כאשר האינדקס הנוסף f בביטוי A_j^f מציין כי מדובר על האלל A_j באוכלוסייה f.

כאשר האינדקס הנוסף f בביטוי $p^f(i)$ מציין כי מדובר על הסתברות למצוא את האלל A_i באוכלוסייה f.

כאשר $P(A_i^u, A_j^u, A_i^v, A_j^v)$ במשוואה (36) מציין את ההסתברות למצוא גם באוכלוסייה u וגם באוכלוסיה v, פרופילים גנטיים להם האללים $\{A_i, A_j\}$ באתר מסויים.

משוואה 16 תוחלף ותהיה:

$$\overline{rmp_v^u}(k) = \sum_{i,j,i \le j}^{n_k} [P(A_i^u, A_j^u, A_i^v, A_j^v)] \quad (37)$$

כאשר ערכי $P(A_i^u, A_j^u, A_i^v, A_j^v)$ נתונים במשוואה (36). באותו אופן, גם משוואה (24) העושה שימוש במשוואה (21) ובמשוואה (8) תוחלף במשוואה (37).

123 שם.

124 ראה בנספח אצל Weir, לעיל ה"ש 114, ביטוי אלגברי המפשט את תוצאת אותה סכימה.

הערות למאמר

הערות כלליות

המאמר מתבסס על עבודה תיאורטית אשר התבצעה על סמך נתוני הדנ"א במאגר העבריינים של מדינה אריזונה ותוצאותיה העלו לכאורה, ספק בנדירות הפרופילים עליהם מדווחות המעבדות הפורנזיות בחוות דעתן. עבודה זאת זכתה לביקורת רבה בתקופה האחרונה[1] וכן לא הצליחה לערער את חוזקה של ראית הדנ"א בבתי המשפט.

קיימים הבדלים עקרוניים ומשמעותיים בין חיפושים תיאורטיים ומודלים תיאורטיים כפי שהם מוצגים במאמר שלפנינו לבין החיפוש הנעשה בפועל במאגר המשטרתי לצורך הפללת אדם. הכותבים מכוונים לעולם המשפט בארץ אך למרבה הצער אינם מסבירים כלל את ההבדלים בין התיאוריה למעשה. עובדה זאת עלולה להכשיל ולהטעות את בתי המשפט בארץ.

בעבודה התיאורטית שנעשתה באריזונה נעשו השוואות חלקיות בין זוגות פרופילים מתוך מאגר הדנ"א המשטרתי עצמו אשר מכיל גם קרובי משפחה ואינו עונה על כל כלל גנטי, ובנוסף ההתאמות שנמצאו לא היו מלאות אלא רק בחלק מהאתרים. חלקיות ההתאמות נקבעו באופן אקראי ולא בהכרח לגבי אתרים מסוימים, כלומר החיפוש היה בדרגת חופש גבוהה, והתוצאה שהתקבלה הייתה מספר האתרים המשותפים שנמצאו, בלי קשר לזהותם. חיפוש בדרגת חופש כזאת לא מתאפשר בחיפוש במאגר המשטרתי. מצב כזה של התאמה חלקית, ב-9 מתוך 10 אתרים בתיק פורנזי, נחשבת כשלילת האפשרות שמקור הפרופיל יכול להיות בחשוד.

הסבר ממצה להבדלים הללו חיוני להיות כלול במאמר מסוג זה[1] ובהיעדרו עשוי בית משפט בארץ לקבל החלטה מוטעית.

לא התייחסתי בבדיקתי למודל התיאורטי ולא לנספח הטכני שכן אינני בקיאה בתחום זה.

הערות מפורטות

עמ' 72: לא ברורה הסקת המסקנה לגבי תיקון טטה כאשר מתבססים על הנתונים מתוך המאגר כיוון שמאגר הדנ"א המצוי בידי המשטרה כולל דגימות על פי סוג העברה שהמחוקק קובע ולא על פי תת אוכלוסיה מסוימת. כמו כן, הדגימות אינן אקראיות ומכילות קרובי משפחה כך שאינן מהוות מודל לגבי מבנה האוכלוסייה בארץ, כולל לגבי תיקון טטה.

פרק ז' עמ' 79 פסקה ראשונה: הפרופיל שהתקבל מתשעת האתרים שנקבעו לדגימה מזירת העבירה ואשר התאים באופן מלא לפרופיל החשוד הינו בשכיחות מאד נמוכה באוכלוסיה. הניסיון להטיל ספק במסקנתו של השופט חשין הינה בגדר הטעיה: על פי ניסוי אריזונה השאלה הייתה אחרת: התבקשו למצוא זוגות אנשים שלהם 9 אתרים משותפים, 9 אתרים כלשהם מתוך 13 אתרים בתוך מאגר של כ-70,000 פרופילים תוך התעלמות מאי התאמה באתרים אחרים.

פרק ז' עמ' 79 פסקה שניה: אכן המשטרה חוקרת יותר מפשע אחד אך לא מיליארדים שזה מספר ההשוואות שנעשו במודל המוצג כאן. כמו כן פרופילים המתאימים על פי הקריטריונים המקובלים בקהילה הפורנזית חסרי דרגת חופש כלל בעוד המודל המוצע הינו בעל דרגת חופש רבה ויכול למצוא התאמות בחלק מהאתרים ולא תמיד באותם אתרים.

עמ' 80 פסקה שניה לפני סוף העמוד: הטיעון כאן איננו ברור. א. אם נמצאה התאמה בין הפרופיל מהזירה ופרופיל החשוד ושכיחות הפרופיל היא אחד ליותר ממיליארד, מדוע יש צורך לבדוק אתרים נוספים. ומדוע דווקא 15 אתרים יתכן ורק 50 אתרים ישביעו את רצון ההגנה? ב. לגבי החיסכון המוצע למערכת, לא ברור מדוע לאחר שנמצא החשוד המתאים, והשכיחות להתאמה אקראית באוכלוסיה מאד נמוכה, יש צורך בבדיקות עשרות חשודים נוספים?

פרק ח' עמ' 81: בפרק זה בו עובר המחבר ממודלים תיאורטיים לטענות מעשיות בנושא שגיאות אפשריות בבדיקות הדנ"א, בתיקי עבודה הנבדקים בארץ, האמירות עלולות <u>להטעות</u> את בתי המשפט.

1 BUDOWLE B. ET AL. *PARTIAL MATCHES IN HETEROGENEOUS OFFENDER DATABASES DO NOT CALL INTO QUESTION THE VALIDITY OF RANDOM MATCH PROBABILITY CALCULATION*. INT J MED 2009 JAN;123(1):59-63

עמ' 81 פסקה שניה: אין דמיון בין ההגדרות של המודל התיאורטי בדבר התאמות חלקיות, לבין פרופילים חלקיים המתקבלים בבדיקות פורנזיות. בתחום התיאורטי, כפי שנאמר קודם לכן, קימת דרגת חופש גדולה לגבי התאמת הפרופילים: לא חשוב אילו אתרים מתאימים ומה אחוז ההתאמה מתוך הפרופיל המלא. בתחום הפורנזי אכן מתקבלים לעיתים, בעיקר מדוגמאות קשות, פרופילים חלקיים (למשל: פרופיל המורכב מ-7 אתרים במקום 10) אך כדי לקבוע התאמה היא חייבת להיות מלאה: זהות מוחלטת באותם האתרים המכילים תוצאות בשני הפרופילים (זה החלקי מזירת העבירה וזה המלא של החשוד). האם יש באוכלוסיה פרופיל זהה נוסף לפרופיל החלקי שנקבע ניתן להעריך לפי השכיחות המחושבת לפרופיל זה שאכן תהיה גבוהה משכיחות של פרופיל בן 10 אתרים אך עדיין פרופיל זה יהיה נדיר באוכלוסיה (במקום אחד למאות מיליארדים רק אחד לכמה מיליארדים).

פסקה שלישית: המשפט השני בפסקה זאת ממש מקומם: פרק זה מכוון לעבודה פורנסית (ראה כותרת) ובהמלצותיו מכוון המחבר להחלטות בית משפט. האמירה "בכל מאגר גדול סביר מאד שיש פרופילים שהופקו באופן שגוי" מטעה. ציטוט זה לקוח מתוך מאמר (הערת שוליים 66 שם) שאינו עוסק כלל בתחום הפורנזי ובנוסף קובע שהשגיאות הינן זניחות. כללי העבודה עליהן מסתמך המאמר הנ"ל שונים לחלוטין . בתחום הפורנזי העוסק בדיני נפשות, כללי העבודה מאד נוקשים ומבוקרים. אמירת הכותבים יוצרת רושם מוטעה אצל הקורא בדבר התכנות גבוהה של שגיאות בקביעת הפרופילים הגנטיים גם בתחום הפורנזי.

הדוגמא לגבי השגיאה בניו-זילנד הינה כשלון נקודתי של מעבדה ולא ניתן להסיק מכישלון נקודתי לגבי הכלל.

המשפט האחרון בפסקה זאת מטעה: אנומליה בהשוואת פרופילים של שני דורות איננה שגיאה, כפי שכותב המחבר, אלא מוטציה אשר המדען הפורנזי מכיר היטב, יודע מה שכיחות הופעתה ולוקח שכיחות זאת בחשבון בהערכתו הסטטיסטית.

עמ' 81 פסקה רביעית: הציטוט (ה"ש 71 שם) הינו של מחבר שאינו עוסק בתחום הפורנסי, שכאמור מחויב בכללים נוקשים משאר התחומים בבואו לקבוע פרופיל גנטי.

יש הטעיה בפירוש הנתונים מציטוט (ה"ש 76 שם): מדובר על BLIND TRIAL שבוצע בין השנים 1993-2003. מאז שנים אלה הטכנולוגיות מאד השתנו לכן התוצאות אינן רלוונטיות. כמו כן מזכיר המחבר את אחוזי השגיאה בשנות האלפיים, כאשר הכוונה בשנים 2000-2003. גם מאז 2003 ועד היום טכנולוגיות רבות השתנו וכללים פורנזיים מחמירים הוכנסו. בנוסף, המחבר טועה כאשר מצטט את אחוז השגיאה המתייחסת להבנתו, לאתר בודד ומכאן מחשב ל-10 אתרים. אותו מבחן עיוור המצוטט (ה"ש 76 שם) מבוצע גם על ידי מעבדתנו ואחוז ההצלחה מתייחס לדגימות נבדקות ולא לאתרים בודדים לכן כל החישוב של המחבר שגוי.

עמ' 82 פסקה ראשונה: תופעת השמטת אלל (Allele Dropout) היא תופעה מדעית ידועה, אשר מופיעה לעיתים כאשר כמות הדנ"א נמוכה ו/או האיכות ירודה (תופעות שכיחות בעולם הפורנזי). כל מעבדה פורנזית מתמודדת עם תופעה זאת על ידי הימנעות מקריאת אלל בודד באתר כאשר הוא מופיע בערכי קריאה הנמוכים מסף מסוים שנקבע על ידי אותה מעבדה. תופעות Null Allele ו-Silent Allele אינן קשורות לתופעת Allele Dropout. בעוד זאת האחרונה לא קיימת בכמויות דנ"א גבוהות, שתי הראשונות לא ייעלמו גם בערכי דנ"א גבוהים.

עמ' 82 פסקה שניה: השוני האפשרי בשימוש בשתי ערכות בעלות רצף פריימרים שונה אינו קשור לנושא ה-Allele Dropout. הנושא מוכר על סמך עבודות וולידציה והמעבדות מתייחסות בהתאם.

עמ' 82 פסקה אחרונה והמשכה בעמ' 83, כולל פסקה שניה: הפסקה מיותרת כיוון שתופעת Allele Dropout אינה בגדר הפתעה או שגיאה וכמובן שהיא עשויה להופיע בתערובת כמו שהיא מופיעה בפרופיל יחיד. כל מעבדה פורנזית מכינה עבודות וולידציה ומכירה היטב מה הוא סף ערכי הקריאה של הפרופיל הגנטי מתחתיו יש חשש לתופעת allele dropout. באזהרות שלא ניתן להבחין בביטחון בין אתר הומוזיגוטי אמיתי ובין אתר הטרוזיגוטי המציג אלל אחד ובו חלה תופעת Allele Dropout, יש משום הטעיה.

עמ' 83 פסקה אחרונה: שגיאה: יצרן ערכת ה-SGM PLUS ממליץ על סף של 50 RFU.[2]

2 Applied Biosystems, GeneMapper ID Software Ver. 3.1, pp 5-20

עמ' 84: המלצה א': למיטב ידיעתי אין היום מעבדה פורנזית בעולם, המיישמת למעשה המלצה זאת. דווקא דוגמאות בעייתיות הן זעירות בכמותן וכמובן שלא ניתן לבצע מספר ריאקציות PCR בודדות אלא ריאקצית multiplex. המלצה ב': המלצה זאת אינה מבחינה בין תופעות שנות: בין תופעה סטוכסטית המתבטאת ב-Allele Dropout או בחוסר איזון בין אללים לבין תופעה של פרופיל שמקורו ביותר מתורם אחד אשר גם היא מתבטאת בחוסר איזון בין אללים. המחבר טועה במסקנותיו לגבי החישובים הנעשים במעבדות בארץ בנושא זה.

ד"ר מאיה פרוינד

מנהלת המעבדה לביולוגיה משפטית
המכון הלאומי לרפואה משפטית
maya.freund@forensic.health.gov.il